Infinite Sets

Lecture notes for NMAI074

Jakub Smolík

smolikj@matfyz.cz

Date of last change: February 4, 2026

Contents

1 Review of Set Theory Basics

We will be working within Zermelo–Fraenkel (ZF) set theory; that is, Zermelo's (1908) theory Z[1] augmented [2] by Fraenkel's (1922) axiom schema of replacement and von Neumann's (1925) axiom of foundation. If we further include the axiom of choice, we obtain the much stronger theory ZFC.

As for notation, I always use the symbol $\subset$ for a proper subset (or subclass) and $\subseteq$ for a general subset (or subclass). I use $\subsetneq$ only when it is important that the two sets (or classes) are not equal. Concatenated expressions such as $a \in b \in c$ mean $a \in b \wedge b \in c$. I differentiate between the symbol for equality of two objects "=" and the symbol for the definition of an object ":=". I use the following notation for defining functions.

- $f : A \to B$ is a function with domain A and codomain B.
- $f : a \mapsto b$ denotes that f maps the set $a \in A$ to the set $b \in B$.
- $f = g \circ h$ means that $f(x) = h(g(x))$ for all suitable x.

I use the terms "function," "map," and "mapping" interchangeably.

1.1 Sets and Classes

Definition 1.1 (Class)**.** If $\varphi(x)$ is a formula, then the expression $\{x \mid \varphi(x)\}$ is called a *class term*. It defines the "collection" of all sets x satisfying $\varphi(x)$. We call this collection the *class* determined by $\varphi(x)$.

Every set is a class, but not all classes are sets (consider the class of all sets). A class that is not a set is called a *proper class*. The major difference between sets and classes is that classes cannot be members of other classes or sets, while sets can. We can substitute class terms into logical formulas in place of free variables, but unlike sets, we cannot quantify them using $\forall$ and $\exists$. It is not hard to show that for every formula with class terms (but without quantified class variables), there is an equivalent formula in the base language without class terms.

We will usually denote sets using small letters $a, b, c, x, y, \ldots$ and classes using capital letters A, B, C, etc. The exception to this are well-ordered sets, which will often be denoted as W. Finally, the class of all sets, also called the *universal class*, is denoted by V.

1.2 Axiom Schema of Replacement

Axiom 1.2. When we take any (even a class) map F and a preimage set a, then the class of images $b = F[a]$ is also a set. Formally, if $\psi(x,y)$ is a formula without free variables y_1, y_2 and b, then

$$(\forall x)(\forall y_1,y_2)\big((\psi(x,y_1) \wedge \psi(x,y_2)) \Rightarrow y_1 = y_2\big) \Rightarrow \\ (\forall a)(\exists b) : (\forall y)\big(y \in b \Leftrightarrow (\exists x)(x \in a \wedge \psi(x,y))\big)$$

[1] Zermelo's theory Z actually already contained AC, which he formulated in 1904.

[2] Replacement is necessary to prove the existence of some key sets, as demonstrated in Section 3.3.3. Foundation is more of a "cleanup" axiom, as virtually all results in the branches of mathematics based on set theory hold even without it.

is an axiom. The formula $\psi(x,y_1)$, resp. $\psi(x,y_2)$ are created from $\psi(x,y)$ by substituting y_1, resp. y_2 for y.

The first part of this axiom says that $\psi(x,y)$ should behave like a map $y = F(x)$. In the second part, a denotes the set of preimages and b the set of corresponding images.

1.3 Axiom of Choice

The *axiom of choice*, denoted AC, is one of the most important principles in modern mathematics, with profound implications in areas such as classical analysis or linear algebra. It states that for any collection of nonempty sets, it is possible to choose exactly one element from each set, even if the collection is infinite. When added to Zermelo–Fraenkel set theory, it yields the much more powerful ZFC. Many theorems that seem intuitively true, such as every vector space having a basis, depend on this axiom.

However, the axiom of choice is also controversial, as it leads to counter-intuitive results, such as the well-ordering principle, which claims that every set can be well-ordered, or the Banach–Tarski paradox, which provides a way to decompose a solid ball into finitely many pieces and reassemble them into two identical copies of the original.

Definition 1.3 (Choice function)**.** A *choice function* (or a *selector*) on the set x is any function $f : x \to \bigcup x$ such that

$$(\forall t \in x)(t \neq \varnothing \Rightarrow f(t) \in t).$$

We can WLOG assume that the choice function is defined on $x \setminus \{\varnothing\}$ and all $t \in \mathrm{Dom}(f)$ satisfy $f(t) \in t$.

One can prove in ZF via finite induction that every finite set has a choice function; that is, we are allowed to make finitely many choices (out of potentially infinite sets). However, it can be shown that ZF cannot prove that every countable set has a choice function, and certainly not that *every* set has a choice function.

Since the assumption that *every* set has a choice function can lead to some paradoxical results (such as the Banach–Tarski paradox), we distinguish three different "power levels" of this axiom.

Axiom 1.4 (Axiom of Countable Choice AC_ω)**.** Every countable set has a choice function; one can make only a countable number of choices.

Axiom 1.5 (Axiom od Dependent Choice DC)**.** One can make a countable sequence of choices, where each choice may *depend* on the previous ones. Formally, for any nonempty set A and a binary relation $R \subseteq A \times A$ such that

$$(\forall x \in A)(\exists y \in A)\, x \, R \, y,$$

there exists an infinite sequence (x_n) satisfying $x_n \, R \, x_{n+1}$ for all $n \in \omega$.

Axiom 1.6 (Axiom of Choice AC)**.** Every set has a choice function; the amount of choices is not limited.

Exercise 1. Show that $\mathsf{AC} \implies \mathsf{DC} \implies \mathsf{AC}_\omega$.

We will now present a brief overview of some of the results that can be proven from each power level (but cannot be proven in ZF). More details and proofs can be found in [10].

Axiom of countable choice AC_ω

$\Longleftrightarrow$ an arbitrary Cartesian product of countably many nonempty sets is nonempty.

$\Longrightarrow$ any union of a countable collection of countable sets is countable.

$\Longrightarrow$ every infinite set has a countably infinite subset. Or equivalently $\omega \preceq x$ for all infinite sets x.

$\Longrightarrow$ every set x is finite $\Longleftrightarrow$ it is Dedekind finite; that is $(\forall y)(y \subset x \Rightarrow y \prec x)$.

$\Longrightarrow$ a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point $x \Longleftrightarrow$ for every convergent sequence (x_n) we have $\lim f(x_n) = f(\lim x_n)$.

Axiom of dependent choice DC

$\Longrightarrow$ a partially ordered set is well-founded (every nonempty subset has a minimal element) $\Longleftrightarrow$ it contains no infinite strictly descending sequences.

$\Longrightarrow$ most results of classical analysis and topology.

Axiom of choice AC

$\Longleftrightarrow$ an arbitrary Cartesian product of nonempty sets is nonempty.

$\Longrightarrow$ there exists a mapping which assigns to each set x a set $|x|$ such that $x \approx |x|$ and $x \approx y \Longleftrightarrow |x| = |y|$.

$\Longleftrightarrow$ for any infinite set x it holds that $|x| = |x \times x|$.

$\Longleftrightarrow$ every vector space (even of infinite dimension) has a basis.

$\Longleftrightarrow$ the product of (even infinitely many) compact topological spaces is compact.

$\Longleftrightarrow$ every surjection $f : X \to Y$ has a right inverse, i.e. a function $g : Y \to X$ such that $f(g(y)) = y$ for all $y \in Y$.

$\Longleftrightarrow$ every connected (even infinite) graph has a spanning tree.

$\Longrightarrow$ the Compactness Theorem in first-order logic: if every finite subset of a theory T has a model, then T has a model.

$\Longrightarrow$ every infinite set has a non-principal ultrafilter.

$\Longleftrightarrow$ **the Well-Ordering Principle**: every set can be well-ordered.

$\Longleftrightarrow$ **Zorn's Lemma**: every ordered set containing upper bounds for every chain necessarily contains at least one maximal element.

$\Longleftrightarrow$ **the Trichotomy Principle**: for any sets x and y either $x \preceq y$, or $y \preceq x$.

We will show the equivalence[3] of AC and the last three conditions in Sections 2.4, 2.5 and 2.6.

[3] "The Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?" — Jerry Bona

Paradoxical results implied by AC

$\Longrightarrow$ the Banach–Tarski paradox: it is possible to decompose a solid ball into a few pieces and reassemble them into two identical copies of the original ball. Vsauce has a VIDEO with an intuitive explanation.

$\Longrightarrow$ there exist subsets of $\mathbb{R}$ that are not Lebesgue measurable. The most famous such set is probably the Vitali set. Veritasium has a great VIDEO on this topic (and the history of AC in general).

The axioms of countable and dependent choice are implicitly used in disciplines such as classical analysis all the time. The full power of the axiom of choice is rarely needed, and we try to avoid it when possible.

1.4 Natural Numbers and the Axiom of Infinity

We use Von Neumann ordinals, meaning that natural numbers are defined as

$$0 \coloneqq \varnothing,\ 1 \coloneqq \{0\},\ 2 \coloneqq \{0,1\}, \dots,\ n+1 \coloneqq \{0,1,\dots,n\} = n \cup \{n\}.$$

Definition 1.7. The *successor function* is a mapping $S : V \to V$ defined as $v \mapsto v \cup \{v\}$. For convenience, we write $v + 1 \coloneqq S(v) = v \cup \{v\}$.

Definition 1.8. A set w is *inductive* if $0 \in w$ and for all $n \in w$ also $n + 1 \in w$.

Axiom 1.9 (Axiom of Infinity)**.** There exists an inductive set.

Definition 1.10. We define the *set of all natural numbers* as the $\subseteq$–smallest inductive set. Or, equivalently, as $\bigcap\{w \mid w \text{ is inductive}\}$. We denote it by ω.

1.5 Comparing Sizes of Sets

Definition 1.11. For sets x and y we define the relations

(a) $x \approx y$, if there exists a bijection $x \to y$, ... x and y are *equinumerous*

(b) $x \preceq y$, if there exists an injection $x \to y$, ... x is *subvalent* to y

(c) $x \prec y$, if $x \preceq y$ and $x \not\approx y$.

Theorem 1.12 (Cantor–Bernstein)**.** $x \approx y \iff (x \preceq y \wedge y \preceq x)$.

Definition 1.13. We say that a set x is

(a) *finite* if $x \approx n$ for some $n \in \omega$, and *infinite* otherwise,

(b) *countable* if $x \preceq \omega$,

(c) *countably infinite* if $x \approx \omega$,

(d) *uncountable* if it is not countable ($x \not\preceq \omega$).

It is easy to see that finite sets are subvalent to infinite sets and that countably infinite and uncountable sets are infinite.

Exercise 2. Verify that a countable set that is not countably infinite is finite.

Hint. Consider what happens if a subset $A \subseteq \omega$ is bounded or unbounded, and use the result of Exercise 5.

Definition 1.14. A set x is *Dedekind-finite* if $y \prec x$ for each $y \subset x$; otherwise, it is *Dedekind-infinite*; that is, there exists an equinumerous subset.

Clearly, each finite set is Dedekind-finite, and thus every Dedekind-infinite set is infinite. However, the other implication does not hold in ZF.

Fact 1.15. *There exist models of* ZF *containing infinite but Dedekind-finite sets.*

Observation 1.16. *If a set x is infinite but Dedekind-finite, then it contains no countably infinite subset $y \subseteq x$; that is, $\omega \not\preceq x$.*

Proof. Suppose x has a countable subset $y = \{a_0, a_1, a_2, \ldots\}$. Then we can construct a bijection $g : x \to x \setminus \{a_0\}$ as $g(a) = a$ if $a \notin y$, and $g(a_n) = a_{n+1}$. $\square$

Hence such sets x are uncountable, but it is not true that $\omega \prec x$.

Theorem 1.17 (AC_ω)**.** *Every infinite set contains a countably infinite subset. Thus every infinite set is Dedekind-infinite, and every Dedekind-finite set is finite.*

Proof. Let x be an infinite set. Because $n \prec x$ for each $n \in \omega$, we can *choose* an injection $G_n : n \to x$ using AC_ω for each $n \in \omega$. Note that each G_n corresponds to a sequence $a_0, a_1, \ldots, a_{n-1}$. Writing the elements of all the sequences $G_0, G_1, G_2, \ldots, G_n, \ldots$ one after another and erasing all occurrences of each member a of x other than its first occurrence yields an infinite sequence H of length ω of different members of x. H is indeed infinite since for every n the sequence G_n consists of n different terms. The formal definition of the sequence H is left to the reader. $\mathrm{Rng}(H)$ is a countably infinite subset of x. $\square$

Corollary 1.18 (AC_ω)**.** *A set x is uncountable $\iff$ $\omega \prec x$*

Proof. "$\Rightarrow$": x is an infinite set, so by the previous theorem $\omega \preceq x$. But $x \not\approx \omega$ since $x \not\preceq \omega$. "$\Leftarrow$": if x were countable, that is $x \preceq \omega$, then by the Cantor–Bernstein theorem (since $\omega \preceq x$) we have $x \approx \omega$, a contradiction with $\omega \prec x$. $\square$

1.6 Well-Orderings and Initial Segments

Definition 1.19 (Ordering)**.** A binary relation R on the class X is a

(a) *trichotomy* if for all x,$y \in X$, either $x = y$, or $x\,R\,y$, or $y\,R\,x$,

(b) *strict order* if it is anti-reflexive, strongly anti-symmetric, and transitive on X; (note that strong anti-symmetry follows from the other two),

(c) *(partial) order* if it is reflexive, weakly anti-symmetric, and transitive on X,

(d) *total (or linear) order* if it is a trichotomous partial order on X.

If R is an ordering, then instead of $x\,R\,y$ we write $x \le_R y$ and we call $(X, \le_R)$ an *ordered class*. Similarly, if R is a strict ordering, then we write $x <_R y$ and we call $(X, <_R)$ a *strictly ordered class.*

Note that we can easily create a strict ordering $<_R$ from $\leq_R$ and vice versa. For this reason, we will not define properties for both strict and non-strict orderings separately, because one implicitly defines the other.

Definition 1.20. We call an element of an ordered class $(X, \leq)$ *minimal* if there is no smaller one, and we call it a *minimum* if it is smaller than all others. If a minimum exists, we denote it by $\min_{\leq}(X)$. The *supremum* of a subset $Y \subseteq X$ is the minimum of all its upper bounds. If it exists, we denote it by $\sup_{\leq}(Y)$

Observation 1.21. *Every minimum is minimal. Furthermore, if $\leq_R$ is a total order, then there is at most one minimal element, and if it exists, then it is also the minimum. There is always at most one minimum.*

Definition 1.22 (Well-ordering)**.** An ordered class $(A, \leq_R)$ is

(a) *well-founded* if every non-empty subset of A has a minimal element.

(b) *well-ordered* if every non-empty subset of A has a minimum (least element).

Notice that every well-ordered class is totally ordered since we can take any two elements, and one of them has to be the minimum and is therefore smaller.

Observation 1.23. *Well-order $\iff$ well-founded total order.*

Observation 1.24. *The well-ordering property is hereditary. That is, if X is well-ordered by $\leq_R$, then every $Y \subseteq X$ is also well-ordered by $\leq_R$.*

Observation 1.25. *Well-founded ordered sets contain no infinite strictly decreasing sequences, as such a sequence has no minimal element.*

Exercise 3. Prove that the reverse implication also holds, provided we accept the axiom of dependent choice (see Axiom 1.5).

Definition 1.26 (Lower part and subset)**.** Let $(A, <_R)$ be a (strictly) ordered class. A subclass $X \subseteq A$ is a *lower part* of A if

$$(\forall x \in X)(\forall a \in A)(a <_R x \Rightarrow a \in X).$$

Additionally, if X is a set, we call it a *lower subset* of A, and if $X \neq A$, then we call it a *proper lower part*, or *proper lower subset* of A.

Lemma 1.27. *Let $(W, <_R)$ be a (strictly) well-ordered set, and suppose that X is a proper lower subset of W. Then there exists a unique $x \in W$ such that X is equal to the set $\{y \in W \mid y <_R x\}$. We denote this set as $(\leftarrow, x)$.*

Proof. We define x as the minimum of $W \setminus X$. Then every $y <_R x$ belongs to X, so $(\leftarrow, x) \subseteq X$. We also want the opposite inclusion. For contradiction, suppose there is a $y \in X$ such that $y \notin (\leftarrow, x)$. If $y \not<_R x$, then necessarily $x <_R y$ as $x \neq y$ since $x \notin X$. But this means that $x \in X$ because X is a lower subset and $y \in X$. But this is a contradiction since $x \notin X$. □

Definition 1.28 (Initial segment)**.** If $(W, <_R)$ is a (strictly) well-ordered set, then we call its proper lower subsets *initial segments* instead. We denote the unique initial segment of W determined by $x \in W$ as

$$(\leftarrow, x) \coloneqq \{y \in W \mid y <_R x\}.$$

It contains all the elements of W from the minimum of W until x, but not x itself.

Observation 1.29. *Note that $x <_R y \iff (\leftarrow, x) \subset (\leftarrow, y)$.*

2 Ordinal Numbers

Informally, *ordinal numbers* are a way to generalize natural numbers. We will first do a quick recap of the basics of ordinal numbers and then prove a theorem that deeply links ordinals and well-ordered sets.

2.1 Ordinals as a Generalization of Naturals

Definition 2.1. A class X is called *transitive* if for all $x \in X$ we have $x \subseteq X$. Or equivalently, if for every x, y such that $y \in x \in X$ we have $y \in X$.

Theorem 2.2. *Every natural number and the set of all natural numbers ω is transitive and (strictly) well-ordered by the membership relation $\in$.*

From now on, we will denote the (strictly) well-ordered set $(\omega, \in)$ as $(\omega, <)$ instead and write $n < m$ instead of $n \in m$ when talking about natural numbers.

Definition 2.3 (Ordinal numbers)**.** A set α is an *ordinal number* if it is transitive and (strictly) well-ordered by the membership relation $\in$. If α is infinite, we say that it is a *transfinite ordinal*. We denote the *class of all ordinal numbers* by On.

Theorem 2.4. *Finite ordinals are exactly the natural numbers, and ω is the smallest transfinite ordinal.*

Theorem 2.5. *The class* On *itself is transitive and (strictly) well-ordered by $\in$. This implies that it is not a set; otherwise,* $\mathrm{On} \in \mathrm{On}$*. Furthermore, any proper class X that is transitive and well-ordered by $\in$ is identical to* On*.*

As for notation, we will usually denote ordinals using letters from the beginning of the Greek alphabet: $\alpha, \beta, \gamma, \delta \ldots$ An exception to this is the letter λ, which we will reserve for limit ordinals. Furthermore, we compare ordinals using the symbol '$<$'. That is, we write $\beta < \alpha$ instead of $\beta \in \alpha$.

Observation 2.6. *If $\beta < \alpha$, then $\beta \subset \alpha$ and β is an initial segment of α. Additionally, $\alpha = (\leftarrow, \alpha)$.*

Definition 2.7. If α is an ordinal, then we call all $\beta < \alpha$ the *predecessors* of α. The *successor* of α is the ordinal $\alpha + 1 := \alpha \cup \{\alpha\}$. We say that α is the *direct predecessor* of $\alpha + 1$.

Remark. It is easy to show that $\alpha + 1$ is the smallest ordinal larger than α.

Definition 2.8. We say that an ordinal number α is an

(a) *isolated* ordinal if $\alpha = 0$ or α has a direct predecessor,

(b) a *limit* ordinal otherwise.

Isolated ordinals $\alpha > 0$ are also sometimes called *successor* ordinals.

Example. Every $n \in \omega$ is isolated, ω is limit, and $\omega + 1$ is isolated again.

2.2 Ordinals as Types of Well-Ordered Sets

The definition of ordinals presented above was formalized by John von Neumann in 1923. This elegant approach, however, came decades after Georg Cantor first introduced ordinals (around 1885) as *order types of well-ordered sets.* Cantor's intuition was that ordinals serve as labels for well-ordered sets: the smallest element is labeled 0, the next 1, and so on. The *order type* of the set is then the first label we did not have to use; it represents the "shape" of the ordering.

Consider, for example, a set ordered as

$$a_0 < a_1 < a_2 < \overbrace{\cdots}^{\infty} < b.$$

Here, there are countably infinitely many elements a_i, followed by one additional element b. If we label the elements from left to right, all the natural numbers are used for the a_i's, leaving no finite label for b. This is precisely why we need transfinite ordinals: we assign the label ω to b. Hence, the order type of this ordering is $\omega + 1$.

It is important to realize that different orderings of the same sets can have different order types. This means that the ordinal numbers do not count the number of objects in the set; they only label them.

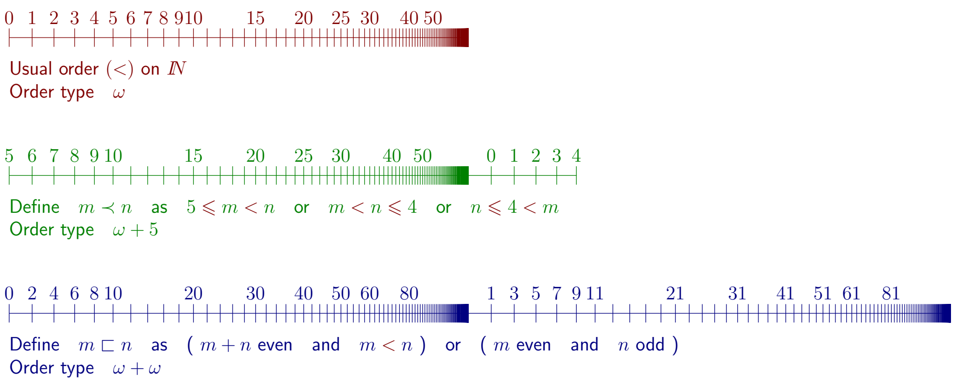

Figure 1: Different orderings of the same set can have different types [5].

Lemma 2.9. *Every proper lower part of* (On, $<$) *is an ordinal number.*

Proof. Let X be a proper lower part of On. Then

(i) X is transitive. Suppose $\alpha \in \beta \in X$, that is $\alpha < \beta \in X$. Because X is a lower part, we have $\alpha \in X$.

(ii) X is well-ordered by $\in$ because On is well-ordered by $\in$ and $X \subseteq \mathrm{On}$.

We also need to argue that X is a set. If it were a proper class, then by Theorem 2.5 it would be the entire On, but $X \subsetneq \mathrm{On}$. □

Definition 2.10 (Isomorphism)**.** Let $(A, \leq_R)$ and $(B, \leq_S)$ be ordered classes. A bijection $F : A \to B$ is an *order-isomorphism* of $(A, \leq_R)$ and $(B, \leq_S)$ if

$$(\forall x,y \in A)\big(x \leq_R y \iff F(x) \leq_S F(y)\big).$$

Because we will not be dealing with other types of isomorphisms, we will usually simply say *isomorphism* instead of order-isomorphism.

Theorem 2.11 (About comparing well-orderings)**.** *If* $(W_1, \leq_1)$ *and* $(W_2, \leq_2)$ *are well-ordered sets, then exactly one of the following holds:*

(a) *either* W_1 *and* W_2 *are isomorphic, or*

(b) W_1 *is isomorphic to an initial segment of* W_2*, or*

(c) W_2 *is isomorphic to an initial segment of* W_1*.*

In each case, the isomorphism is unique.

Corollary 2.12. *No two distinct ordinal numbers can be isomorphic.*

Proof. Suppose $\alpha < \beta$, that is $\alpha \in \beta$ and $\alpha \subset \beta$. Clearly α is an initial segment of β. This means that we are in case (b) of the previous theorem. $\square$

Theorem 2.13 (About the type of well-ordering)**.** *Every well-ordered set* W *is isomorphic to a unique ordinal number* α*, which is called the* order type *of* W *and is denoted by* $\mathrm{otp}(W)$*.*

Proof (following [15, Thm. 3.1, Chap. 6]*).* Let $(W, <_R)$ be a well-ordered set. We want to show that there is a unique ordinal $(\alpha, <)$ isomorphic to it. Define X as the set of all $x \in W$ for which $(\leftarrow, x)$ is isomorphic to an ordinal. As no two distinct ordinals are isomorphic, this ordinal is uniquely determined, and we denote it α_x; we denote the isomorphism as $i_x : (\leftarrow, x) \to \alpha_x$.

Suppose that there exists a set S such that $S = \{\alpha_x \,|\, x \in X\} \subseteq \mathrm{On}$. Because we assume that S is a set, then $S \subsetneq \mathrm{On}$. We claim that S is a proper lower part of $(\mathrm{On}, <)$, and thus, by Lemma 2.9, it is an ordinal; let's call it α. Indeed, suppose $\beta < \alpha_x \in S$, we want $\beta \in S$. Note that β is an initial segment of α_x. This implies that $i_x^{-1}[\beta]$ is an initial segment of W. Because W is well-ordered, $i_x^{-1}[\beta]$ is equal to $(\leftarrow, b)$ for some $b \in W$ (using Lemma 1.27). So $\beta = \alpha_b \in S$ by the definition of S. More precisely, $i_x \restriction (\leftarrow, b)$ is an isomorphism between $(\leftarrow, b)$ and β. We will argue why we can make the assumption that S is a set later.

A similar argument shows that X is a lower subset of W. To show this, suppose $x \in X$ and take $y \in W$ such that $y <_R x$. We want $y \in X$. We have $y <_R x$; therefore, $(\leftarrow, y)$ is an initial segment of $(\leftarrow, x)$. Because isomorphisms conserve all ordering properties, $i_x \restriction (\leftarrow, y)$ is an isomorphism between $(\leftarrow, y)$ and an initial segment of α_x. By Lemma 2.9, this is an ordinal; by our previous notation, α_y. Therefore $y \in X$.

We conclude that either $X = W$ or $X = (\leftarrow, c) \subset W$ for some $c \in W$ (using Lemma 1.27). We now define a function $f : X \to S = \alpha$ by $f : x \mapsto \alpha_x$. From the definition of S and the fact that

$$x <_R y \iff (\leftarrow, x) \subset (\leftarrow, y) \iff \alpha_x \subset \alpha_y \iff \alpha_x < \alpha_y,$$

it is obvious that f is an isomorphism of $(X, <_R)$ and $(\alpha, <)$. If

- $X = (\leftarrow, c)$, then by the definition of the set X, $c \in X$ because $(\leftarrow, c)$ is isomorphic to an ordinal $\alpha_c = \alpha$. But this is a contradiction because $c \notin (\leftarrow, c) = X$.

- Therefore, $X = W$ and α is the sought-after ordinal isomorphic to $(W, <_R)$.

The uniqueness of α follows from the simple observation that if W were isomorphic to two distinct α_1 and α_2, then by the transitivity of isomorphisms, the ordinals α_1 and α_2 would be isomorphic, which is impossible by Corollary 2.12.

This would complete the proof if we were justified to make the assumption that the class S is a set and therefore an ordinal. In fact, we have to use the axiom of replacement to guarantee it. If we assume this axiom, then S is a set because it is the image of the set X under the map f. $\square$

Exercise 4. Is there a well-ordered proper class not isomorphic to $(\mathrm{On}, <)$?

Hint. Try to modify On to contradict the property described in Lemma 2.9. If you run out of ideas, consult Section 3.2.

2.3 Transfinite Induction and Recursion

In mathematics, we often use induction on the natural numbers to prove statements, and we can use recursion, such as $f(0) = 1$ and $f(n) = n \cdot f(n-1)$, to define functions. We will now show how to generalize this to all ordinals.

Theorem 2.14 (Transfinite Induction Principle)**.** *Let $A \subseteq \mathrm{On}$ be a class such that for all ordinals $\alpha \in \mathrm{On}$, we have $\alpha \subseteq A \Rightarrow \alpha \in A$, or in other words*

$$(\forall \beta < \alpha)(\beta \in A) \implies (\alpha \in A). \tag{2.1}$$

Then $A = \mathrm{On}$.

Equivalently, assume that $\varphi(x)$ is a property, and for all ordinals α:

If $\varphi(\beta)$ holds for all $\beta < \alpha$, then $\varphi(\alpha)$.

Then $\varphi(\alpha)$ holds for all ordinals α.

Proof. Suppose that $\gamma \in \mathrm{On} \setminus A$ and let $S = \{\alpha \leq \gamma \mid \alpha \notin A\}$. Because ordinals are well-ordered, the set S has a minimum element α. Since every $\beta < \alpha$ is in A, it follows by (2.1) that $\alpha \in A$, which is a contradiction.

The equivalence can be easily seen by taking the class $A = \{x \mid \varphi(x)\}$ or the property $\varphi(x) = x \in A$. $\square$

We can also formulate the principle separately for isolated and limit ordinals, which allows us to use the transfinite induction principle in a form closer to the usual formulation of the induction principle for the naturals.

Theorem 2.15 (Transfinite Induction Principle II)**.** *Let $A \subseteq \mathrm{On}$ be a class satisfying*

(i) $0 \in A$,

(ii) $\alpha \in A \Rightarrow \alpha + 1 \in A$, *...this is just induction on ω*

(iii) *if α is a limit ordinal and $(\forall \beta < \alpha)(\beta \in A)$, then $\alpha \in A$.*

Then $A = \mathrm{On}$. Note that we can again easily reformulate this in terms of a property $\varphi(x)$.

Proof. We need to show that these three assumptions imply (2.1). So let α be an ordinal such that $\beta \in A$ for all $\beta < \alpha$. If $\alpha = 0$, then $\alpha \in A$ by (i). If $\alpha \neq 0$ is isolated, that is if there is a $\beta < \alpha$ such that $\alpha = \beta + 1$, we know that $\beta \in A$, so $\alpha \in A$ by (ii). If α is a limit ordinal, we have $\alpha \in A$ by (iii). $\square$

We can use transfinite induction to prove properties of certain infinite structures. On the other hand, transfinite recursion—the technique described in the following theorem—allows us to construct various infinitely complex structures and define functions in a recurrent fashion.

Theorem 2.16 (About construction by transfinite recursion)**.** *If $G : V \to V$ is a class map, then there is a unique class map $F : \mathrm{On} \to V$ satisfying*

$$F(\alpha) = G(F \restriction \alpha). \tag{2.2}$$

So we define the image of the next ordinal using its predecessors and their images.

Remark. This should seem a bit suspicious because it looks like we are saying that for every class G, there exists a class F for which something holds. But we cannot quantify classes. Well, we can replace the quantification of G with a theorem *schema*, one for each G. And we aren't really quantifying F because the following proof explicitly constructs it.

Remark. The generality of defining $F(\alpha)$ based on the ordered pairs $(\beta, F(\beta))$ for all $\beta < \alpha$ allows us to define functions using many other recursions. For example:

- $F(\alpha) = G\big(F[\alpha]\big) = G\big(\{F(\beta) \mid \beta < \alpha\}\big)$,
- $G : \mathrm{On} \times V \to V$ and $F(\alpha) = G(\alpha, F \restriction \alpha)$,
- $F(\alpha)$ is $G_1(F(\beta))$ if $\alpha = \beta + 1$ is isolated, and $G_2(F[\alpha])$ if α is limit. This is the form we will use most often.

Additionally, the theorem about construction by transfinite recursion is equivalent to the axiom of replacement, as shown in [14].

Proof. We define A as the class of "set approximations" of F. That is set mappings f, the domain of which is some ordinal number β, and that for all $\alpha < \beta$, we have $f(\alpha) = G(f \restriction \alpha)$. Now we define F as $F := \bigcup A$. Clearly $F \subseteq \mathrm{On} \times V$. We will show that $F : \mathrm{On} \to V$ is the unique mapping satisfying (2.2).

First, we show that the approximations of F agree. Let $f, f' \in A$ and $\alpha \in \mathrm{Dom}(f) \cap \mathrm{Dom}(f')$. We claim that $f(\alpha) = f'(\alpha)$. Note that $\mathrm{Dom}(f) \cap \mathrm{Dom}(f')$ is an ordinal δ. For contradiction, suppose that $\alpha \in \delta$ is the smallest ordinal for which $f(\alpha) \neq f'(\alpha)$. Then $f \restriction \alpha = f' \restriction \alpha$ so $f(\alpha) = G(f \restriction \alpha) = G(f' \restriction \alpha) = f'(\alpha)$, a contradiction.

Second, we verify that F satisfies (2.2); that is, for all $\alpha \in \mathrm{Dom}(F)$, we have $F(\alpha) = G(F \restriction \alpha)$. So let $\alpha \in \mathrm{Dom}(F)$. It is there due to some $f \in A$ satisfying $\alpha \in \mathrm{Dom}(f)$ and $f(\alpha) = G(f \restriction \alpha)$. Also, $F(\alpha) = f(\alpha)$ and $F \restriction \alpha = f \restriction \alpha$. Therefore, by combining these equalities $F(\alpha) = G(F \restriction \alpha)$.

Next, we show that $\mathrm{Dom}(F) = \mathrm{On}$. First, we prove that $\mathrm{Dom}(F)$ is a lower part of On. Suppose $\alpha \in \mathrm{Dom}(F)$; then it is there thanks to some $f \in A$ with domain $\delta > \alpha$. If $\beta < \alpha$, then also $\beta \in \delta$, and thus $\beta \in \mathrm{Dom}(F)$.

According to Lemma 2.9, either $\mathrm{Dom}(F) = \mathrm{On}$, which we want, or $\mathrm{Dom}(F) = \gamma \in \mathrm{On}$. Suppose, for contradiction, that $\mathrm{Dom}(F) = \gamma$. Then F is a set because $\mathrm{Dom}(F)$ is a set, $\mathrm{Rng}(F)$ is a set using the axiom of replacement, and $F \subseteq \mathrm{Dom}(f) \times \mathrm{Rng}(f)$. This implies that $F \in A$ because its domain is an ordinal, and we have verified that it satisfies the recursive definition property.

Now that $F \in A$, we define a slightly "longer" function $F_1 \coloneqq F \cup \{(\gamma, G(F))\}$; note that $F = F_1 \restriction \gamma$. Notice that $F_1 \in A$ because $\mathrm{Dom}(F_1) = \gamma + 1$ is an ordinal, and we defined it to satisfy the recursive definition property. Because $F = \bigcup A$, this implies $F_1 \subseteq F$, but then $\gamma \in \mathrm{Dom}(F_1) \subseteq \mathrm{Dom}(F) = \gamma$, which is a contradiction. We conclude that $\mathrm{Dom}(F) = \mathrm{On}$.

Finally, we prove the uniqueness of F. For contradiction, suppose that there is another mapping $F' \neq F$ satisfying this theorem. Because $(\mathrm{On}, <)$ is well-ordered, we can take the smallest ordinal α where $F(\alpha) \neq F'(\alpha)$. Therefore $F \restriction \alpha = F' \restriction \alpha$ and so $F(\alpha) = G(F \restriction \alpha) = G(F' \restriction \alpha) = F'(\alpha)$, which is a contradiction. $\square$

Exercise 5. Prove by induction on ω that every infinite well-ordered set A, such that each initial segment $(\leftarrow, a)$ is finite, is isomorphic to $(\omega, <)$.

Hint. Since each $(\leftarrow, a)$ is finite, there is a unique $n_a \in \omega$ with the same cardinality. The isomorphism we are looking for is $f : A \to \omega$ defined by $f : a \mapsto n_a$.

Exercise 6. Prove by transfinite recursion that every well-ordered proper class W, such that each proper lower part $(\leftarrow, a)$ is a set, is isomorphic to $(\mathrm{On}, <)$.

Hint. Use transfinite recursion to define an order-isomorphism $F : \mathrm{On} \to W$ using $G(x) = \min(W \setminus x)$ and $F(\alpha) = G(F[\alpha])$. Then $F(0) = G(\varnothing) = \min(W)$, and $F(\alpha)$ is smallest element of W that has not yet been used to define $F(\beta)$ for some smaller $\beta < \alpha$.

We will use transfinite recursion to prove the equivalence of the well-ordering theorem and Zorn's lemma to the axiom of choice. But transfinite recursion can also be used to prove some wildly sounding geometrical claims, such as

- $\mathbb{R}^3$ is a union of pair-wise disjoint unit circles, or that
- there is a set in $\mathbb{R}^2$ that intersects every line in exactly two points.

2.4 Well-Ordering Principle

The *well-ordering principle*—the statement that every set can be well-ordered—was a foundational belief of Georg Cantor, but he was unable to provide a proof for it. This challenge was famously solved by Ernst Zermelo in 1904. Zermelo was the first person to explicitly state the axiom of choice, which he identified as the principle Cantor (and many others) had been implicitly using in many proofs. He then demonstrated that AC and the well-ordering principle are equivalent, which is why the principle is now often called the "Well-Ordering Theorem" or "Zermelo's Theorem." Veritasium has a great video [26] on this topic.

Principle 2.17 (Well-Ordering Principle)**.** Every set can be well-ordered.

Theorem 2.18. *The well-ordering principle is equivalent to the axiom of choice.*

Proof. $\mathsf{WO} \Rightarrow \mathsf{AC}$. Let $A \neq \varnothing$ be a set, without loss of generality $\varnothing \notin A$. We want to construct a selector $f : A \to \bigcup A$ such that for all $a \in A$ we have $f(a) \in a$. The well-ordering principle guarantees a well-ordering $\leq$ on $\bigcup A$, and because every a is a nonempty subset of $\bigcup A$, it has a least element with respect to $\leq$. We chose this minimum as $f(a)$.

$\mathsf{AC} \Rightarrow \mathsf{WO}$. Let $A \neq \varnothing$ be a set. We will use transfinite recursion to label the elements of A by ordinal numbers and then use the well-order of the ordinals to define a well-order on A. Let $g : \mathcal{P}(A) \to A$ be a selector on $\mathcal{P}(A)$, assigning to each nonempty $B \subseteq A$ an element $b \in B$. We will want to use transfinite recursion based on g, so we should extend it to be a class map $G : V \to V$, for example, by defining it to be equal to $\varnothing$ when g is not defined.

We can now use transfinite recursion to define the function $F : \mathrm{On} \to A \cup \{\varnothing\}$ as $F(0) = G(A)$ and $F(\alpha) = G(A \setminus F[\alpha])$. This function assigns to each ordinal a unique element from A until they "run out" (when $F[\alpha] = A$), and then it assigns $\varnothing$ to all larger ordinals.

Define W as the class of all ordinals α for which $F[\alpha] \subsetneq A$. Denote the restriction of F to W as $F_W : W \to A$. Plan: first, we show that W itself is an ordinal. From this, it will follow that F_W is a bijection between W and A, allowing us to denote the unique ordinal mapped to $a \in A$ as α_a. Once this is established, we define a well-ordering R of A as

$$a <_R b \iff \alpha_a < \alpha_b.$$

This is a well-ordering since $(A, <_R)$ is order-isomorphic to $(W, <)$, which is well-ordered (as W is an ordinal).

Firstly, we claim that W is a set. Indeed, because F_W is injective, it has an inverse F_W^{-1} that maps the set $\mathrm{Rng}(F_W) \subseteq A$ onto W, which is therefore, using the axiom of replacement, a set. Now we claim that W is a lower subset of On, and so it is an ordinal (by Lemma 2.9). Suppose $\alpha \in W$, that is $F[\alpha] \subsetneq A$, and let $\beta < \alpha$. Then $\beta \subseteq \alpha$ and $F[\beta] \subseteq F[\alpha]$, so $\beta \in W$.

To complete the proof, we must show that $F_W : W \to A$ is a bijection. It is clearly injective. To show that it is surjective, suppose for contradiction that there exists some $b \in A \setminus F_W[W]$. Because W is an ordinal number γ, it satisfies the definition of W (thanks to b) and thus $W = \gamma \in W$, which is a contradiction. $\square$

2.5 Zorn's Lemma

Zorn's lemma is perhaps the most useful application of the axiom of choice outside of set theory. It is also known as the maximality principle, a name that dates back to the German mathematician Felix Hausdorff, who proved an earlier and equivalent version of the theorem in 1914 (see [29] for details). The formulation known today as Zorn's lemma was introduced in 1935 by the German mathematician Max Zorn. However, it had already been independently proved in 1922 by the Polish mathematician Kazimierz Kuratowski, whom you might know for Kuratowski's theorem—a forbidden-graph characterization of planar graphs.

Definition 2.19 (Chain)**.** Let $(a, \leq_R)$ be an ordered set. We call a subset $b \subseteq a$ a *chain* if b is totally ordered by $\leq_R$.

Principle 2.20 (Zorn's Lemma)**.** Every (partially) ordered set containing upper bounds for every chain necessarily contains at least one maximal element.

There is also a parameterized version of this statement.

Principle 2.21 (Parametrized Zorn's Lemma)**.** Let A be a (partially) ordered set containing upper bounds for every chain. Then for every $a \in A$, there is a maximal element $b \in A$ such that $a \leq b$.

We can obtain the parameterized version from the unparameterized one by restricting ourselves to the elements above or equal to a. The other direction is obvious.

Remark. Zorn's lemma can be made slightly stronger by assuming that only well-ordered chains have upper bounds. The proof remains virtually unchanged.

Theorem 2.22. *The axiom of choice implies Zorn's lemma.*

Proof. Let $(A, <_R)$ be an ordered set containing upper bounds for each chain and for contradiction suppose that there is no maximal element. Note that this implies that every chain, in fact, has a *strict* upper bound. If a chain C had no strict upper bound, then the non-strict upper bound $b \in C$ would be a maximal element. We denote the set of strict upper bounds of C as $C^>$.

We take $f : \mathcal{P}(A) \to A$, a selector on $\mathcal{P}(A)$, and define a function g from the set of all chains in A as $g(C) \coloneqq f(C^>)$. So g maps a chain to one of its strict upper bounds. Now pick an arbitrary $a \in A$ and define the mapping $H : \mathrm{On} \to A$ by transfinite recursion as $H(0) = a$ and $H(\alpha + 1) = g(\{H(\alpha)\})$ for successor ordinals, and as $H(\delta) = g(H[\delta])$ for limit ordinals. We start with a and get larger and larger elements of A using successor ordinals, each time taking a strict upper bound of a single element chain. If an ordinal δ is limit, we notice that $H[\delta]$ is a chain (all the smaller elements that we picked previously are strict upper bounds of each other and are therefore comparable), and $H(\delta)$ is a strict upper bound of this chain.

Note that if we want to be rigorous about the construction by transfinite recursion, we should define g on the entire V. But we can do this in any way, for example, by defining $G(x)$ as $\varnothing$ if x is not a chain of A, and $g(x)$ otherwise.

Finally, observe that $H : \mathrm{On} \to A$ is an increasing function (each value is a strictly larger upper bound than the previous one) and that it is injective. Thus, we obtain an injection from the proper class On into the set A, which is impossible. Indeed, taking the inverse mapping and applying the axiom of replacement would imply that On itself is a set, which is a contradiction. □

Theorem 2.23. *Zorn's lemma implies the well-ordering principle.*

Proof. Let X be any set. We will find a well-ordering of it by considering all of its possible well-ordered subsets, picking the maximal one using Zorn's lemma, and showing that it orders the entire X. Consider the set: [4]

$$\mathcal{W} \coloneqq \{(A, <_R) \,|\, <_R \text{ is a well-order on } A \subseteq X\},$$

and define a partial order $\prec_{\mathcal{W}}$ on it by $(A, <_R) \prec_{\mathcal{W}} (B, <_S)$ if B end-extends A. That is, if $A \subset B$, and $<_R$ is the restriction of $<_S$ to A, and A is an initial segment of B. We will apply Zorn's lemma to $\mathcal{W}$.

[4] Why is this a set?

First, we need to show that chains have upper bounds. Let $\mathcal{C} \subseteq \mathcal{W}$ be a chain. Define the set

$$M \coloneqq \bigcup \{A \mid (A, <_R) \in \mathcal{C}\} \subseteq X,$$

and for $x, y \in M$ put $x <_M y$ if there exists some $(A, <_R) \in \mathcal{C}$ such that $x, y \in A$ and $x <_R y$. Because $\mathcal{C}$ is a chain, this is well-defined: if x and y belong to two distinct orderings in $\mathcal{C}$, then one extends the other and hence they agree.

We claim that $(M, <_M)$ is well-ordered. Let $S \subseteq M$ be nonempty and pick some $s \in S$. Then $s \in A_s$ for some $(A_s, <_R) \in \mathcal{C}$. Note that $A_s \cap S$ is nonempty, and because A_s is well-ordered, there exists a minimum $m = \min_{<_R}(A_s \cap S)$. Notice that also $m = \min_{<_M}(S)$. Indeed, if there were a $t \in S \setminus A_s$ such that $t <_M m$, then it would be in S due to some $A_t \in \mathcal{C}$ containing t. Since both A_s and A_t are in the chain, either

(a) $A_t \subseteq A_s$, which is impossible since then $t \in A_s$, or

(b) $A_s \subset A_t$, meaning that A_s is an initial segment of A_t, and therefore $m \in A_s$ is smaller than $t \in A_t \setminus A_s$, which contradicts the assumption that $t <_M m$.

Therefore $(M, <_M)$ is well-ordered and thus an upper bound of $\mathcal{C}$ in $\mathcal{W}$.

Because all chains are bounded, by Zorn's lemma, $\mathcal{W}$ has a maximal element $(W, <_W)$. We claim that $W = X$ and so it is the sought-after well-ordering of X. For contradiction, suppose there exists some $x \in X \setminus W$ and extend the ordering $<_W$ to $W' \coloneqq W \cup \{x\}$ by making each $y \in W$ smaller than x. Notice that this slightly "longer" order is a well-ordering of W' and therefore is in $\mathcal{W}$. Moreover, it end-extends $(W, <_W)$ which hence is not maximal in $(\mathcal{W}, \prec_{\mathcal{W}})$. We have arrived at a contradiction and can conclude that $W = X$. □

Exercise 7. Would the proof still have worked if instead of end-extensions, we had simply used general extensions? Meaning that the smaller ordering doesn't need to be an initial segment of the larger one.

Hint. By defining the end-extension ordering, we have ensured that chains have a similar structure to chains of ordinals (larger ordinals end-extend the smaller ones). Thus, when proving that M is well-ordered, we could have used a similar strategy as when proving that the ordinals are well-ordered.

To demonstrate an application of Zorn's lemma, consider the following question. Does every connected graph have a spanning tree? Finding one in a finite graph is easy: simply remove the edges of cycles until there are no cycles left. But this process may not terminate for infinite graphs.

Proposition 2.24. *Every connected graph has a spanning tree.*

Sketch of proof. The set of all sub-graphs that are trees is partially ordered by inclusion, and the union of a chain is its upper bound. Zorn's lemma states that a maximal tree must exist, which is a spanning tree since the graph is connected. □

Remark. In general, suppose that we have a structure represented by a set X (a graph) with substructures $A \subseteq X$ (subgraphs that are trees), and we want to show that there is a maximal substructure. Then we simply need to check that the union of a chain of substructures is itself a substructure.

2.6 Trichotomy Principle

Principle 2.25 (Trichotomy principle)**.** The relation $\preceq$ is is trichotomous on V. That is, for any sets x and y either $x \preceq y$, or $y \preceq x$.

Trichotomy seems only natural: how "weird" would sets x and y really have to be to be completely incomparable? We have mentioned in Section 1.5 that there exist models of ZF containing infinite but Dedekind-finite sets, and that such sets have no countably infinite subset. Therefore, if we let x be such a set, then $\omega \npreceq x$, but also clearly $x \npreceq \omega$, otherwise x would be Dedekind-infinite.

Theorem 2.26. *Zorn's lemma implies the trichotomy principle.*

Proof. Let x, y be arbitrary sets; we want to find an injection $x \to y$ or $y \to x$. Consider the set[5]

$$\mathcal{F} = \{f \mid f \text{ is an injection, } \mathrm{Dom}(f) \subseteq x \text{ and } \mathrm{Rng}(f) \subseteq y\}.$$

Notice that the ordered set $(\mathcal{F}, \subseteq)$ satisfies the conditions of Zorn's lemma since the union of a chain of injections is again an injection. Let g be a maximal element of $\mathcal{F}$. If both $x \setminus \mathrm{Dom}(g)$ and $y \setminus \mathrm{Rng}(g)$ were non-empty, then it would be possible to extend g by an extra pair, contradicting its maximality. Hence either $\mathrm{Dom}(g) = x$ and then $x \preceq y$, or $\mathrm{Rng}(g) = y$ and then $y \preceq x$. Here, we used the fact that the inverse of an injection is also an injection. $\square$

Later, (Theorem 3.48), we will show that the trichotomy principle implies the well-ordering principle and thus also the axiom of choice.

Theorem 2.27. *We conclude that the following statements are equivalent in* ZF*:*

(1) *the axiom of choice,*

(2) *the well-ordering principle,*

(3) *Zorn's lemma,*

(4) *the trichotomy principle.*

3 Operations on Ordinals

3.1 Ordinal Functions

Definition 3.1 (Ordinal function)**.** We say that a mapping F is an *ordinal function* if its domain is a lower part of On, that is $\mathrm{Dom}(F) \in \mathrm{On}$ or $\mathrm{Dom}(F) = \mathrm{On}$, and $\mathrm{Rng}(F) \subseteq \mathrm{On}$. We say that F is

(a) *increasing* if for all $\beta \in \mathrm{Dom}(F)$ and $\alpha < \beta$ we have $F(\alpha) < F(\beta)$, and

(b) *non-decreasing* if for all $\beta \in \mathrm{Dom}(F)$ and $\alpha < \beta$ we have $F(\alpha) \leq F(\beta)$.

Remark. We do not define decreasing ordinal functions as they would not be very interesting — the well-ordering of On does not allow infinite decreasing sequences (Observation 1.25), hence F can be decreasing only when $\mathrm{Dom}(F)$ is finite.

[5] Why is this a set?

Lemma 3.2. *Increasing ordinal functions grow at least as fast as the identity function. That is $F(\alpha) \geq \alpha$ for every $\alpha \in \mathrm{Dom}(F)$ for increasing F.*

Proof. For contradiction, suppose that α is the least ordinal such that $F(\alpha) < \alpha$. This means that for every $\beta < \alpha$, we have $F(\beta) \geq \beta$ (note that $\beta \in \mathrm{Dom}(F)$). Suppose $\beta = F(\alpha)$; then $F(\beta) \geq \beta = F(\alpha)$, which is a contradiction since F is increasing. $\square$

Lemma 3.3. *If α and β are the ordinal types of the well-ordered sets A and $B \subseteq A$, then $\beta \leq \alpha$. In other words $B \subseteq A \implies \mathrm{otp}(B) \leq \mathrm{otp}(A)$.*

Note that $B \subset A$ does not imply that $\beta < \alpha$; consider ω and $\omega \setminus \{\varnothing\}$.

Proof. Let $i_a : A \to \alpha$ and $i_b : B \to \beta$ be the isomorphisms of A and B with their types. Suppose $\beta > \alpha$ and define $f : \beta \to \alpha$ as $f = i_b^{-1} \circ i_a$. Notice that if $\gamma < \delta$, then $f(\gamma) < f(\delta)$ because both i_b^{-1} and i_a preserve order (they are order-isomorphisms), and thus f is increasing. Because $\alpha \in \mathrm{Dom}(f)$, we have that $f(\alpha) \in \mathrm{Rng}(f) \subseteq \alpha$, so $f(\alpha) < \alpha$. But this contradicts the previous lemma. $\square$

Recall what you know about metric spaces, namely about closed sets and continuous functions. A subset X of a metric space M is closed if for each convergent sequence $(a_n) \subset X$ we have that $\lim a_n \in X$.

You might also recall that a function f is continuous $\iff$ the preimage $f^{-1}[Y]$ of every closed set Y is closed $\iff$ for every sequence (a_n) we have that $f(\lim a_n) = \lim f(a_n)$. The first equivalence statement is utilized in topology to define continuous functions, and we could use it here as well. However, the second equivalence seems more natural, since $\lambda = \sup\{\alpha \mid \alpha < \lambda\}$ for any limit ordinal λ. Hence limit ordinals essentially represent limits of sequences.

Lemma 3.4. *If $A \subseteq \mathrm{On}$ is a set, then $\bigcup A \in \mathrm{On}$, and in fact, $\bigcup A = \sup(A)$. We say that $\sup(A)$ is the* limit *of the sequence of ordinals A.*

Remark. We use the term "sequence" informally here. What we really mean is that we can label the elements of A as $A = \{\alpha_\delta \mid \delta < \gamma\}$ for some ordinal γ. We use $(\alpha_\delta)_{\delta<\gamma}$ to denote the bijection $\delta \mapsto \alpha_\delta$.

Definition 3.5 (Closed class)**.** A subclass $C \subseteq \mathrm{On}$ is *closed* if, for every subset $Y \subseteq C$, we have $\sup(Y) \in C$. For an ordinal α, we say that a subset $C \subseteq \alpha$ is *closed in α* if, for every $Y \subseteq C$ satisfying $\sup(Y) < \alpha$, we have $\sup(Y) \in C$.

Observation 3.6. *If C is a closed set, then it has a maximum* $\max(C) = \sup(C)$*.*

Example. The ordinal ω is not closed as $\sup(\omega) = \omega \notin \omega$, but $\omega + 1$ is closed as $\sup(\omega + 1) = \omega \in \omega + 1$, and for any $y \subset \omega + 1$ we have $\sup(y) \leq \sup(\omega + 1)$.

Observation 3.7. *An ordinal number α is closed $\iff$ it is isolated.*

Definition 3.8 (Normal function)**.** An ordinal function F is *continuous* if for every limit ordinal $\lambda \in \mathrm{Dom}(F)$ it holds that

$$F(\lambda) = \sup\{F(\alpha) \mid \alpha < \lambda\}.$$

We say that a function is *normal* if it is increasing and continuous.

Example. The simplest normal function is identity. But consider the (very innocent looking) function $F(\alpha) = \alpha + 1$. It is increasing, but not continuous. It fails on limit ordinals, for example $F(\omega) = \omega + 1$, but

$$\sup\{F(\alpha) \mid \alpha < \omega\} = \sup\{\alpha + 1 \mid \alpha < \omega\} = \sup(\omega \setminus \{\varnothing\}) = \omega.$$

Observation 3.9. *If F is an increasing ordinal function and λ is a limit ordinal, then $\sup\{F(\alpha) \mid \alpha < \lambda\}$ is also a limit ordinal. Specifically, if F is normal and λ limit, then $F(\lambda)$ is limit as well.*

Proof. Let $A = \{F(\alpha) \mid \alpha < \lambda\}$ and suppose that $\sup(A)$ is isolated. Notice that if A is a set of ordinals and $\sup(A)$ is isolated, then $\sup(A) \in A$, so $\sup(A) = \max(A)$. Hence, there exists some $\alpha < \lambda$ such that $F(\alpha) = \max(A)$. But this is a contradiction because F is increasing, so $F(\alpha + 1) > F(\alpha)$. □

Exercise 8. Show that the *topological definition of continuity* would make sense. Prove that an increasing function F is continuous $\iff$ the preimage $F^{-1}[C]$ of every closed set $C \subset \mathrm{On}$ is closed in $\mathrm{Dom}(F)$.

Observation 3.10. *The composition $F \circ G$ of normal functions F and G is a normal function. This can be seen easily from the previous exercise.*

Lemma 3.11. *If F is a normal function, then for every ordinal β, such that $F(0) \leq \beta < \sup \mathrm{Rng}(F)$, the maximum $\max\{\alpha \mid F(\alpha) \leq \beta\}$ exists.*

Intuition. For a natural number β and $F(n) = n^2$, we might consider the largest natural number α such that $F(\alpha) \leq \beta$. This α exists, it is in fact equal to $\lfloor\sqrt{\beta}\rfloor$.

Proof. Notice that the set $[0, \beta] := \{\alpha \mid \alpha \leq \beta\}$ is closed by Observation 3.7 because $[0, \beta] = \beta + 1$ is an isolated ordinal. We will use the topological definition of continuity (Exercise 8) and note that the preimage C of the closed set $[0, \beta]$ is closed in $\mathrm{Dom}(F)$. We would like to say that C is closed (in general). But consider $F : \omega \to \mathrm{On}$; then $\mathrm{Dom}(F)$ is not closed in On.

The bound on β will save us. Notice that there is some $\gamma \in \mathrm{Rng}(F)$ such that $\beta < \gamma \leq \sup \mathrm{Rng}(F)$. Because F is increasing, the elements of C are bounded by $F^{-1}(\gamma)$, and therefore $\sup(C) \in \mathrm{Dom}(F)$. Since C is closed in $\mathrm{Dom}(F)$, it follows that $\sup(C) \in C$ and $\sup(C) = \max(C)$, which we will denote as α. Because F is increasing, α is the largest ordinal satisfying $F(\alpha) \leq \beta$. □

Lemma 3.12. *If $K \subseteq \mathrm{On}$ is a closed proper class, then there exists a unique bijective normal ordinal function $J : \mathrm{On} \to K$ enumerating the elements of K. Equivalently we can say that J is a continuous order-isomorphism of* On *and K.*

Proof. The proper class K inherits from On its well-order and the property that every proper lower part $(\leftarrow, a) \subset K$ is a set (see Lemma 2.9). Exercise 6 claims that there is a unique isomorphism (bijective increasing function) $J : \mathrm{On} \to K$.

To show that J is continuous, let λ be a limit ordinal. Because it is the first ordinal larger than all $\alpha < \lambda$, it will be mapped to the first $\kappa \in K$ larger than all $J(\alpha)$ for $\alpha < \lambda$. Because K is closed, $\delta := \sup\{J(\alpha) \mid \alpha < \lambda\} \in K$. We claim that $\kappa = \delta$ (hence J is continuous). If not, then δ was already used by some $\beta < \lambda$, that is $J(\beta) = \delta$. Because J is increasing, we have for $\beta + 1 < \lambda$ that $J(\beta + 1) > J(\beta) = \delta$, so δ is not the supremum, a contradiction. □

Definition 3.13 (Fixed point)**.** We call an ordinal ξ a *fixed point* F if $F(\xi) = \xi$.

Theorem 3.14 (About fixed points)**.** *Let $F : \mathrm{On} \to \mathrm{On}$ be a normal function.*

(i) *For every $\alpha \in \mathrm{On}$, there exists $\beta \geq \alpha$ that is a fixed point of F.*

(ii) *The first such fixed point is the limit (supremum) of the sequence $(\alpha_n)_{n<\omega}$ defined as $\alpha_0 = \alpha$ and $\alpha_{n+1} = F(\alpha_n)$.*

(iii) *The class K of all fixed points of F is a closed proper class.*

(iv) *There exists a unique bijective normal function $F' : \mathrm{On} \to K$ called the* derivative *of F, which enumerates the fixed points of F.*

Proof. First notice that (iv) is a direct consequence of (iii) by Lemma 3.12. To prove (i) and (ii) notice that $\alpha_{n+1} \geq \alpha_n$ since F grows at least as fast as the identity function, and that the supremum $\beta = \sup\{\alpha_n \mid n \in \omega\}$ is a fixed point:

- Consider the case when $\alpha_0 < \alpha_1 < \cdots < \alpha_i = \alpha_{i+1}$ for some i; then also $\alpha_{i+2} = F(\alpha_{i+1}) = F(\alpha_i) = \alpha_i$ and by induction $\alpha_n = \alpha_i$ for all $n \geq i$, thus $\beta = \alpha_i$ is a fixed point.

- Suppose the sequence never stabilizes. Since F is continuous, we have

$$F(\beta) = \sup\{F(\gamma) \mid \gamma < \beta\} = \sup\{F(\alpha_n) \mid n \in \omega\} = \sup\{\alpha_{n+1} \mid n \in \omega\} = \beta.$$

Second, we show that β is the smallest fixed point larger than α. If there were a fixed point $\xi \geq \alpha$ such that $\xi < \beta$, then there would exist an index n at which $\alpha_n \leq \xi < \alpha_{n+1}$. This is because the sequence is strictly increasing, and β is its supremum. We have $\xi < \alpha_{n+1} = F(\alpha_n) \leq F(\xi)$, so ξ is not a fixed point.

Finally, we prove (iii). We claim that K is closed. Let $C \subseteq K$ be a set; we need to show that the supremum $\beta = \sup(C)$ is a fixed point. Since F is continuous, we have

$$F(\beta) = \sup\{F(\gamma) \mid \gamma < \beta\} = \sup\{F(\xi) \mid \xi \in C\} = \sup\{\xi \mid \xi \in C\} = \beta.$$

To complete the proof, we show that K is a proper class. If it were a set, then by Lemma 3.4, $\sup(K)$ is an ordinal γ. We let the ordinal $\gamma + 1$ take the role of α in (i) and find a new fixed point of F, larger than all those in K. $\square$

Theorem 3.15 (About simultaneous fixed points)**.** *Let $\langle F_i \mid i \in I \rangle$ be a collection of normal functions $F_i : \mathrm{On} \to \mathrm{On}$ indexed by a set I. We say that ξ is a* simultaneous fixed point *of $\langle F_i \mid i \in I \rangle$ if $F_i(\xi) = \xi$ for all $i \in I$.*

(i) *For every $\alpha \in \mathrm{On}$, there exists $\beta \geq \alpha$ that is a simultaneous fixed point of $\langle F_i \mid i \in I \rangle$.*

(ii) *The first simultaneous fixed point $\beta \geq \alpha$ is the limit of the sequence $(\alpha_n)_{n<\omega}$ defined as $\alpha_0 = \alpha$ and $\alpha_{n+1} = \sup\{F_i(\alpha_n) \mid i \in I\}$.*

(iii) *The class K of all simultaneous fixed points is a closed proper class.*

(iv) *There exists a unique bijective normal function $J : \mathrm{On} \to K$ enumerating the simultaneous fixed points of $\langle F_i \mid i \in I \rangle$.*

Proof. Define an ordinal function $\overline{F} : \mathrm{On} \to \mathrm{On}$ as $\overline{F}(\alpha) \coloneqq \sup\{F_i(\alpha) \mid i \in I\}$. This is well defined since we can use replacement to guarantee that $\{F_i(\alpha) \mid i \in I\}$ is a set. Notice that ξ is a fixed point of $\overline{F}$ if and only if it is a simultaneous fixed point of $\langle F_i \mid i \in I \rangle$ since

$$\overline{F}(\xi) = \xi \implies (\forall i \in I) F_i(\xi) \leq \xi \iff (\forall i \in I) F_i(\xi) = \xi \implies \overline{F}(\xi) = \xi.$$

The equivalence holds because all F_i grow at least as fast as the identity function. Note that this also implies that $\overline{F}$ itself grows at least as fast as the identity function. Moreover, $\overline{F}$ is continuous. Indeed, if λ is a limit ordinal, then

$$\begin{aligned} \overline{F}(\lambda) &= \sup\{F_i(\lambda) \mid i \in I\} \\ &= \sup\{\sup\{F_i(\delta) \mid \delta < \lambda\} \mid i \in I\} && \ldots \text{each } F_i \text{ is continuous} \\ &= \sup\{\sup\{F_i(\delta) \mid i \in I\} \mid \delta < \lambda\} \\ &= \sup\{\overline{F}(\delta) \mid \delta < \lambda\}. \end{aligned}$$

In general, $\overline{F}$ is not normal, as it is not guaranteed to be increasing. However, notice that the proof of Theorem 3.14 only uses the fact that F is continuous and that it grows at least as fast as the identity. Hence, we can use the theorem on $\overline{F}$ to obtain the claims (i)–(iv). □

3.2 Ordinal Arithmetic

In this section we define operations such as ordinal addition and multiplication. Before proceeding further, I highly recommend watching the video [27] by Vsauce, which illustrates the concepts of constructing larger ordinals from earlier ones in a very illustrative and intuitive way.

3.2.1 Definitions and Intuition

Definition 3.16. Let α and β be ordinals. We define ordinal numbers

(a) $\alpha + \beta$ as the order type of the set $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$ when ordered lexicographically,

(b) $\alpha \cdot \beta$ as the order type of the set $\beta \times \alpha$ when ordered lexicographically.

Using the popular "matchstick" representation of ordinals, $\alpha+\beta$ can be imagined as a pile of decreasing matchsticks labeled by α, followed by another pile of matchsticks labeled by β. Notice that our previous notation of denoting $\alpha \cup \{\alpha\}$ by $\alpha + 1$ is consistent with the above definition. We first use the elements of α to label the first pile, and we need one additional ordinal to label the second pile (which contains only a single matchstick).

Notice that we are using $\beta \times \alpha$ in the definition of $\alpha \cdot \beta$. The ordinal $\alpha \cdot \beta$ can be imagined as taking multiple piles of matchsticks labeled by α and arranging them next to each other. How should the piles be arranged? In a way that we need β to label them.

With this intuition, it should not be surprising that ordinal addition and multiplication are generally not commutative. It is easy to see that $1 + \omega = \omega$ (label the first pile by 0 and the other pile by $\omega \setminus \{0\}$), but $\omega + 1 \neq \omega$. For multiplication, consider $2 \cdot \omega$, the order type of countably infinitely many copies of $\{0, 1\}$ stacked behind each other. This can be clearly labeled by ω, so $2 \cdot \omega = \omega$. But $\omega \cdot 2$ is the order type of two consecutive copies of ω. When we try to label them using ω, we use all $n \in \omega$ to label the first copy and need more ordinals for the second copy. Therefore $\omega \cdot 2 > \omega$.

Observation 3.17. *For any ordinals α, β, γ and natural $n \in \omega$ it holds that*

(a) $\alpha + 0 = \alpha = 0 + \alpha, \qquad \alpha \cdot 0 = 0 = 0 \cdot \alpha, \qquad \alpha \cdot 1 = \alpha = 1 \cdot \alpha,$

(b) $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma, \qquad \alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma,$

(c) $\alpha \cdot 2 = \alpha + \alpha, \qquad \alpha \cdot 3 = \alpha + \alpha + \alpha, \qquad \alpha \cdot (n + 1) = \alpha \cdot n + \alpha.$

Definition 3.18. For ordinal numbers α and β, we define α^β recursively as

(i) $\alpha^0 := 1$,

(ii) if $\beta = \gamma + 1$ is isolated, then $\alpha^\beta := \alpha^\gamma \cdot \alpha$,

(iii) if β is a limit ordinal, then $\alpha^\beta := \sup\{\alpha^\gamma \mid 0 < \gamma < \beta\}$.

Remark. Formally, for every fixed α we use transfinite recursion to define an ordinal function $F_\alpha : \beta \mapsto \alpha^\beta$.

To get an intuition for ordinal powers, consider the ordinal $\omega^2 = \omega \cdot \omega$. It represents multiple copies of ω arranged in the same manner as ω.

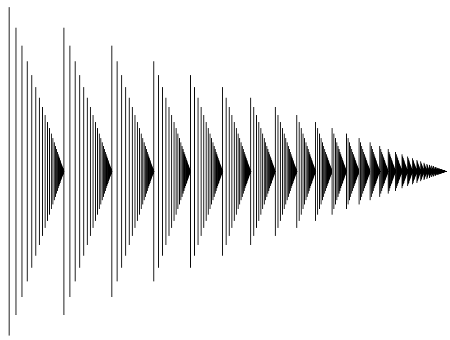

Figure 2: A representation of the ordinal ω^2. Each stick corresponds to an ordinal of the form $\omega \cdot m + n$ where m and n are natural numbers [13].

To construct $\omega^3 = (\omega \cdot \omega) \cdot \omega$, we take multiple copies of ω^2 and arrange them in a way that requires ω to label them. If we repeat this process countably infinitely many times, we arrive at ω^ω.

We can continue and arrive at larger and larger ordinals, such as $\omega^{(\omega^\omega)}$ or $\omega^{(\omega^{(\omega^\omega)})}$. That is a lot of parentheses, so from now on, we will write ω^{ω^ω} instead of $\omega^{(\omega^\omega)}$ and use parentheses only when we mean to say $(\omega^\omega)^\omega$.

Observation 3.19. *For any ordinals α and $\beta > 0$ it holds that*

(a) $0^0 = 1, \quad 0^\beta = 0,$

(b) $1^0 = 1, \quad 1^\beta = 1,$

(c) $\alpha^0 = 1, \quad \alpha^1 = \alpha, \quad \alpha^2 = \alpha \cdot \alpha, \quad \alpha^3 = (\alpha \cdot \alpha) \cdot \alpha.$

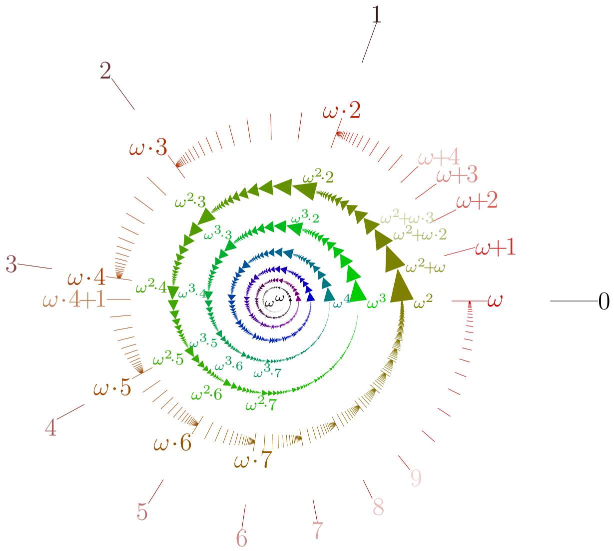

Figure 3: A spiral representation of ordinals up to ω^ω. One full turn corresponds to the mapping $f(\alpha) = \omega \cdot (1 + \alpha)$. Since ω^ω is the smallest fixed point of f, larger ordinals cannot be represented in this way [4].

3.2.2 Properties of Ordinal Operations

You should now have an intuition for how ordinal numbers constructed using these standard operations look. We continue by proving some of their basic properties.

Lemma 3.20 (Monotonicity of sum). *For any α, β and γ it holds that*

(a) $\alpha < \beta \implies \gamma + \alpha < \gamma + \beta$,

(b) $\alpha < \beta \implies \alpha + \gamma \leq \beta + \gamma$.

Proof. (a) From the definition of addition and order types, it is easy to see that $\gamma + \alpha$ is an initial segment of $\gamma + \beta$. (b) The set of ordered pairs that defines $\alpha + \gamma$ is a subset of the set of ordered pairs that defines $\beta + \gamma$. And Lemma 3.3 states that the order type (under lexicographic order) of the first is at most that of the second. ☐

Lemma 3.21 (Monotonicity of product). *For any α, β and $\gamma > 0$ it holds that*

(a) $\alpha < \beta \implies \gamma \cdot \alpha < \gamma \cdot \beta$,

(b) $\alpha < \beta \implies \alpha \cdot \gamma \leq \beta \cdot \gamma$. *. . . for $\gamma = 0$ also holds*

Proof. (a) If $\alpha < \beta$, then $\alpha \times \gamma$ is an initial segment of $\beta \times \gamma$ when ordered lexicographically. (b) If $\alpha < \beta$, then $\gamma \times \alpha \subseteq \gamma \times \beta$, and the claim follows from Lemma 3.3. ☐

Remark. Note that the second statement in the two preceding lemmas does not, in general, hold under strict inequality. For example, $1 < 2$, but

$$1 + \omega = 2 + \omega = \omega, \qquad \text{and} \qquad 1 \cdot \omega = 2 \cdot \omega = \omega.$$

In fact, for any natural $n \in \omega$, we have that $n + \omega = \omega$ and $n \cdot \omega = \omega$.

Lemma 3.22 (Distributivity)**.** *For any ordinals α and β_1, β_2, we have*

$$\alpha \cdot (\beta_1 + \beta_2) = \alpha \cdot \beta_1 + \alpha \cdot \beta_2.$$

That is, ordinal addition and multiplication are left-distributive. However, in general, they are not right-distributive. Meaning that for some α and β_1, β_2

$$(\beta_1 + \beta_2) \cdot \alpha \neq \beta_1 \cdot \alpha + \beta_2 \cdot \alpha$$

Proof. Left distributivity essentially states that if we arrange $\beta_1 + \beta_2$ copies of α next to each other, then it is the same as first arranging β_1 copies of α, followed by β_2 copies of α. This is obviously true from how we defined addition and multiplication. However, in general, these operations are not right-distributive. Consider $(1+1) \cdot \omega \neq \omega + \omega$. On the left, we have $2 \cdot \omega$, and on the right, $\omega \cdot 2$. □

Theorem 3.23. *If m, n and k are natural numbers, then $m + n$, $m \cdot n$, and m^n are also natural numbers. Furthermore*

$$m + n = n + m, \qquad m \cdot n = n \cdot m, \qquad (m + n) \cdot k = m \cdot k + n \cdot k.$$

That is: addition and multiplication of natural numbers is commutative and right-distributive.

Proof. It is easy to see that $m + n$ and $m \cdot n$ are finite ordinals, and one can use induction on n to show that m^n is also finite.

To show that $m \cdot n = n \cdot m$, take any bijection $f : n \times m \to m \times n$ and use it along with the lexicographic order on $n \times m$ to define an isomorphic linear order $<_f$ on $m \times n$. It is well known (and it is proved in the basic set theory course) that any two linear orders on a finite set are isomorphic, so $<_f$ is isomorphic to the lexicographic order of $m \times n$. Since isomorphisms are transitive, the two lexicographically ordered sets $n \times m$ and $m \times n$ are isomorphic and thus have the same order types. Hence $m \cdot n = n \cdot m$. One can similarly show that addition commutes as well.

Right-distributivity is implied by commutativity and left-distributivity. □

Lemma 3.24 (Existence of the right difference)**.** *If $\alpha \leq \beta$, then there is a unique ordinal ϱ such that $\alpha + \varrho = \beta$. We denote ϱ by $\beta \dot{-} \alpha$.*

Intuition. Any ordinal can be extended by a specific amount to reach any larger ordinal. Furthermore, when restricted to natural numbers, $a \dot{-} b$ is the standard subtraction operation.

Proof. If $\alpha \leq \beta$, then $\alpha = (\leftarrow, \alpha)$ is an initial segment of β, and its complement, $\beta \setminus \alpha$, is what we might denote as $[\alpha, \to)$. If ϱ is the order type of $\beta \setminus \alpha$, then clearly $\alpha + \varrho = \beta$. The uniqueness of ϱ follows from Lemma 3.20 (a): suppose there were $\varrho_1 < \varrho_2$ satisfying $\alpha + \varrho_1 = \beta = \alpha + \varrho_2$. But since $\varrho_1 < \varrho_2$, we have that $\alpha + \varrho_1 < \alpha + \varrho_2$. □

Lemma 3.25 (Division with remainder)**.** *If $\beta > 0$, then for every ordinal α there are unique ordinals $\delta \leq \alpha$ and $\varrho < \beta$ such that $\alpha = \beta \cdot \delta + \varrho$.*

Intuition. Any ordinal α can be created by arranging multiple copies of β in a specific way, and following this with a short tail ϱ.

Proof. Since $\beta \geq 1$, we have $\alpha \leq \beta \cdot \alpha$. If $\alpha = \beta \cdot \alpha$ (for example $\omega = 3 \cdot \omega$), choose $\delta := \alpha$ and $\varrho := 0$. It is not hard to show that the monotonicity of sum and product, together with left-distributivity, implies uniqueness; we will skip it.

If $\alpha < \beta \cdot \alpha$, let j be the isomorphism of the lexicographically ordered set $\alpha \times \beta$ and the ordinal $\beta \cdot \alpha$. Let $(\delta, \varrho) \in \alpha \times \beta$ (so $\delta < \alpha$ and $\varrho < \beta$) be the (unique) pair mapped by j to $\alpha \in \beta \cdot \alpha$. Since $\alpha \times \beta$ is ordered lexicographically, it is easy to see that $\alpha = \beta \cdot \delta + \varrho$. $\square$

Lemma 3.26 (Monotonicity of power)**.** *For any α, β, γ and $\rho > 1$, it holds that*

(a) $\alpha < \beta \implies \alpha^\gamma \leq \beta^\gamma$,

(b) $\alpha < \beta \implies \rho^\alpha < \rho^\beta$.

Proof. (a) Using transfinite induction on γ. If $\gamma = 0$, then $\alpha^\gamma = \beta^\gamma = 1$. If $\gamma = \delta + 1$ and $\alpha^\delta \leq \beta^\delta$, from the monotonicity of product we have that

$$\alpha^\gamma = \alpha^\delta \cdot \alpha \leq \beta^\delta \cdot \beta = \beta^\gamma.$$

If γ is a limit ordinal and for every $\delta < \gamma$ already $\alpha^\delta \leq \beta^\delta$, then also

$$\alpha^\gamma = \sup\{\alpha^\delta \,|\, 0 < \delta < \gamma\} \leq \sup\{\beta^\delta \,|\, 0 < \delta < \gamma\} = \beta^\gamma.$$

(b) Suppose that $\rho > 1$. It is easy to show using transfinite induction on δ that for every $\delta > 1$ it holds that $\rho^\alpha < \rho^{\alpha+\delta}$. If $\alpha < \beta$, then according to Lemma 3.24 there is a unique $\delta > 0$ satisfying $\beta = \alpha + \delta$. $\square$

Remark. Note that the first statement in the previous lemma does not, in general, hold under the strict inequality, even if $\gamma > 0$. For example, $2 < 3$, but $2^\omega = 3^\omega = \omega$. In general, if $n \in \omega$, then $n^\omega = \omega$.

Lemma 3.27 (Continuity in the second argument)**.** *The ordinal function*

(a) $F(\xi) = \alpha + \xi$ *is normal for every* $\alpha \geq 0$,

(b) $F(\xi) = \alpha \cdot \xi$ *is normal for every* $\alpha > 0$,

(c) $F(\xi) = \alpha^\xi$ *is normal for every* $\alpha > 1$.

Proof. All of the functions mentioned above are increasing for the specified α, since the respective operations are monotonic. We claim that they are also continuous; that is $F(\lambda) = \sup\{F(\xi) \,|\, \xi < \lambda\}$ for limit λ. Notice that this holds for (c) as this is simply the definition of α^λ.

Because F is increasing, $F(\lambda)$ is an upper bound of the values $F(\xi)$ for $\xi < \lambda$. Assume for contradiction that $F(\lambda)$ is not the smallest upper bound, meaning that there exists $\sigma < F(\lambda)$ such that $\sigma \geq F(\xi)$ for all $\xi < \lambda$.

(a) Lemma 3.24 claims that there is a unique ϱ such that $\sigma = \alpha + \varrho = F(\varrho)$. Because $\sigma < F(\lambda)$ we have that $\alpha + \varrho < \alpha + \lambda$ thus by monotonicity $\varrho < \lambda$ and also $\varrho + 1 < \lambda$. It should hold that $\sigma \geq F(\varrho + 1)$, but $\sigma = F(\varrho) < F(\varrho + 1)$.

(b) Lemma 3.25 claims the existence of unique ordinals δ and $\varrho < \alpha$ such that $\sigma = \alpha \cdot \delta + \varrho$. Since $\sigma < \alpha \cdot \lambda$ we have from monotonicity that $\delta < \lambda$ and also $\delta + 1 < \lambda$. It should hold that $\sigma \geq F(\delta + 1)$, but from monotonicity we have

$$\sigma = \alpha \cdot \delta + \varrho < \alpha \cdot \delta + \alpha = \alpha \cdot (\delta + 1) = F(\delta + 1). \qquad \square$$

Lemma 3.28 (Properties of exponents)**.** *For any α, β and γ, it holds that*

(a) $\alpha^{\beta+\gamma} = \alpha^{\beta} \cdot \alpha^{\gamma}$,

(b) $(\alpha^{\beta})^{\gamma} = \alpha^{\beta\cdot\gamma}$

Proof. (a) Trivially holds for $\alpha \leq 1$. Suppose $\alpha > 1$, we will use transfinite induction on γ. If $\gamma = 0$, there is nothing to prove. If $\gamma = \delta + 1$ is isolated, then $\beta + \gamma = (\beta + \delta) + 1$, and the statement follows from the induction hypothesis:

$$\alpha^{(\beta+\delta)+1} = \alpha^{(\beta+\delta)} \cdot \alpha^{1} = \alpha^{\beta} \cdot \alpha^{\delta} \cdot \alpha = \alpha^{\beta} \cdot \alpha^{\delta+1} = \alpha^{\beta} \cdot \alpha^{\gamma}.$$

Finally, if γ is a limit ordinal and the statement holds for all $\delta < \gamma$, then

$$\begin{aligned}
\alpha^{\beta} \cdot \alpha^{\gamma} &= \sup\{\alpha^{\beta} \cdot \xi \mid \xi < \alpha^{\gamma}\} && \ldots F(\xi) = \alpha^{\beta} \cdot \xi \text{ is normal} \\
&= \sup\{\alpha^{\beta} \cdot \alpha^{\delta} \mid 0 < \delta < \gamma\} && \ldots \alpha^{\gamma} = \sup\{\alpha^{\delta} \mid 0 < \delta < \gamma\} \\
&= \sup\{\alpha^{\beta+\delta} \mid 0 < \delta < \gamma\} && \ldots \text{induction hypothesis} \\
&= \sup\{\alpha^{\varepsilon} \mid 0 < \varepsilon < \beta + \gamma\} && \ldots F(\xi) = \beta + \xi \text{ is normal} \\
&= \alpha^{\beta+\gamma}.
\end{aligned}$$

The last equality holds because $\beta + \gamma$ is a limit ordinal from Observation 3.9.

(b) Suppose that $\beta, \gamma \neq 0$ and $\alpha > 1$, otherwise, it trivially holds. We will again use transfinite induction on γ. If $\gamma = 0$, then it holds. If $\gamma = \delta + 1$ is isolated, then the statement follows from (a) and the induction hypothesis:

$$(\alpha^{\beta})^{\delta+1} = (\alpha^{\beta})^{\delta} \cdot (\alpha^{\beta})^{1} = \alpha^{\beta\cdot\delta} \cdot \alpha^{\beta} = \alpha^{\beta\cdot\delta+\beta} = \alpha^{\beta\cdot(\delta+1)} = \alpha^{\beta\cdot\gamma}.$$

Finally, if γ is a limit ordinal, then $\beta \cdot \gamma$ is also limit and

$$\begin{aligned}
(\alpha^{\beta})^{\gamma} &= \sup\{(\alpha^{\beta})^{\delta} \mid 0 < \delta < \gamma\} && \ldots \gamma \text{ is limit} \\
&= \sup\{\alpha^{\beta\cdot\delta} \mid 0 < \delta < \gamma\} && \ldots \text{induction hypothesis} \\
&= \sup\{\alpha^{\varepsilon} \mid 0 < \varepsilon < \beta \cdot \gamma\} && \ldots F(\xi) = \beta \cdot \xi \text{ is normal} \\
&= \alpha^{\beta\cdot\gamma}.
\end{aligned}$$

The last equality holds because $\beta \cdot \gamma$ is a limit ordinal. □

3.2.3 Ordinal Equations and Power Expansions

Example. Suppose we want to find all ξ and β satisfying $\xi + \beta = \omega$. Lemma 3.24 claims that $\xi \leq \omega$ and $\beta = \omega \dot{-} \xi$. Suppose $\xi = \omega$, then $\beta = 0$. If $\xi = n$ is a natural number, then $\beta = \omega \dot{-} n = \omega$. We conclude that β can attain only two different values.

Proposition 3.29. *Let α be an ordinal and consider the equation $\xi + \beta = \alpha$. The set of solutions (ξ, β) contains only finitely many distinct values of β.*

Proof. Suppose that for some α, there are infinitely many distinct values of β in the solution set. Let $(\xi_n, \beta_n)_{n\in\omega}$ be a sequence of solutions such that $\beta_n < \beta_{n+1}$ for all n. Since $\xi_n + \beta_n = \xi_{n+1} + \beta_{n+1}$, from the monotonicity of sum we have that $\xi_n > \xi_{n+1}$ for all $n \in \omega$. We have constructed an infinite strictly decreasing sequence, which is impossible since On is well-ordered. □

We are able to express any natural number n as an expansion of powers of any base $b > 1$. We will prove that a similar statement holds for ordinal numbers too. A base of special importance is ω (as it is the first transfinite ordinal), and the expansion of α over ω is called its *Cantor normal form*; however, an expansion is possible over any base $\beta > 1$.

Lemma 3.30. *If $k, m_0, m_1, \ldots, m_k$ are natural numbers and $\gamma_0, \gamma_1, \ldots, \gamma_k > \delta$ are ordinals, then*

$$\omega^\delta > \omega^{\gamma_0} \cdot m_0 + \omega^{\gamma_1} \cdot m_1 + \cdots + \omega^{\gamma_k} \cdot m_k.$$

Proof. Let m be the largest among all m_i, and γ be the largest among all γ_i. Then $\omega^\gamma \cdot m \cdot k$ is an upper bound of the sum on the right side of the equation. We assumed that $\delta \geq \gamma + 1$, so $\omega^\delta \geq \omega^{\gamma+1} > \omega^\gamma \cdot m \cdot k$. □

Theorem 3.31 (Expansion over ω)**.** *For any $\alpha > 0$ there are unique natural numbers $k, m_0, m_1, \ldots, m_k \neq 0$ and ordinals $\gamma_0 > \gamma_1 > \cdots > \gamma_k$ which satisfy*

$$\alpha = \omega^{\gamma_0} \cdot m_0 + \omega^{\gamma_1} \cdot m_1 + \cdots + \omega^{\gamma_k} \cdot m_k. \tag{3.1}$$

The sum on the right side of the equation is called the Cantor normal form of α.

Furthermore, if

$$\beta = \omega^{\delta_0} \cdot n_0 + \omega^{\delta_1} \cdot n_1 + \cdots + \omega^{\delta_l} \cdot n_l \tag{3.2}$$

is the Cantor normal form of an ordinal β, then $\beta > \alpha$ if and only if one of the two following cases occurs:

(a) *$l > k$ and the first k terms of β are idential to those of α. For example: $\alpha = \omega^2 + \omega \cdot 2$ and $\beta = \omega^2 + \omega \cdot 2 + 3$.*

(b) *there exists an index $i \leq \min(k, l)$ at which (γ_i, m_i) and (δ_i, n_i) differ, and for the smallest such index i either $\delta_i > \gamma_i$, or $\delta_i = \gamma_i$ and $n_i > m_i$. For example $\alpha = \omega^2 + \omega \cdot 2$ and $\beta = \omega^2 \cdot 5 + \omega \cdot 2$.*

Proof. We prove the first part by transfinite induction on α. The CNF of $\alpha = 1$ is $\alpha = \omega^0 \cdot 1$. Suppose $\alpha > 1$ and that every nonzero $\beta < \alpha$ has a unique CNF. The ordinal function $\gamma \mapsto \omega^\gamma$ is normal, so according to Lemma 3.11, there exists a maximal ordinal γ such that $\omega^\gamma \leq \alpha$. Similarly, from the normality of product in the second argument follows the existence of a maximal ordinal δ such that $\omega^\gamma \cdot \delta \leq \alpha$. Also, $\delta < \omega$, since $\omega^{\gamma+1} = \omega^\gamma \cdot \omega > \alpha$, which contradicts the choice of γ. If $\omega^\gamma \cdot \delta = \alpha$, then the uniqueness of this expansion follows from Lemma 3.30.

If $\omega^\gamma \cdot \delta < \alpha$, then there exists a unique ordinal $\beta = \alpha \dot{-} \omega^\gamma \cdot \delta$ such that $\omega^\gamma \cdot \delta + \beta = \alpha$. Note that $\beta < \omega^\gamma$; otherwise, we get $\omega^\gamma \cdot \delta + \beta \geq \omega^\gamma \cdot (\delta + 1)$, which contradicts the choice of δ. To find the CNF of α, let

$$\beta = \omega^{\gamma_1} \cdot m_1 + \omega^{\gamma_2} \cdot m_2 + \cdots + \omega^{\gamma_k} \cdot m_k$$

be the CNF of β. Define $\gamma_0 \coloneqq \gamma$ and $m_0 \coloneqq \delta$. Then $\gamma_0 > \gamma_1$, and (3.1) is the CNF of α. The uniqueness of this expansion follows from Lemma 3.30 and the unique choice of β.

Next, we prove the second part of the theorem. Suppose that the ordinals α and β have Cantor normal forms (3.1) and (3.2). If (a) holds, then $\beta > \alpha$ because the trailing terms in the expansion of β are nonzero. Suppose that (b) holds and that i is the least index at which the two expansions differ. If $\delta_i > \gamma_i$, then $\beta > \alpha$ from Lemma 3.30. If $\delta_i = \gamma_i$ and $n_i > m_i$, then $n_i \geq m_i + 1$ and

$$\omega^{\delta_i} \cdot n_i \geq \omega^{\gamma_i} \cdot m_i + \omega^{\gamma_i}.$$

Lemma 3.30 claims that the second summand on the right (ω^{γ_i}) is a strict upper bound of the remaining summands in (3.1), the expansion of α; thus $\beta > \alpha$.

All that remains is to prove the reverse implication. If $\beta > \alpha$, then their Cantor normal forms (3.1) and (3.2) have to differ. Either the CNF of one of the ordinals is the same as the beginning of the CNF of the other, or there exists an index at which they differ. We can use the already proven implication to show that the only two possible cases are (a) and (b). □

Corollary 3.32 (Alternative expansions)**.** *For any $\alpha > 0$, it holds that*

(a) *there is a unique natural number $l > 0$ and unique ordinals $\gamma_0 \geq \gamma_1 \geq \cdots \geq \gamma_l$ which satisfy*

$$\alpha = \omega^{\gamma_0} + \omega^{\gamma_1} + \cdots + \omega^{\gamma_l},$$

(b) *there are unique ordinals β and γ such that*

$$\alpha = \omega^{\gamma} \cdot (\beta + 1).$$

Proof. (a) For any ordinal γ and natural m, is the ordinal number $\omega^{\gamma} \cdot m$ equal to the sum of m summands of the form ω^{γ}. We obtain the expansion in (a) by expressing each term in the CNF of α in this expanded form.

(b) If α has CNF (3.1), we let $\gamma = \gamma_k$. Then for all $i \leq k$ is $\gamma_i = \gamma + \delta_i$ for $\delta_i = \gamma_i \dot{-} \gamma$. From the properties of exponents and left-distributivity, we get

$$\alpha = \omega^{\gamma} \cdot \big(\omega^{\delta_0} \cdot m_0 + \omega^{\delta_1} \cdot m_1 + \cdots + \omega^0 \cdot m_k\big).$$

The parentheses on the right contain an isolated ordinal $\beta + 1$, because m_k is a nonzero natural number. The uniqueness of the ordinals γ and β follows from the uniqueness of the CNF of α. □

Theorem 3.33 (Expansion over any base)**.** *The choice of ω as a base in Theorem 3.31 was arbitrary; the same holds for any ordinal base $\beta > 1$. We just need to restrict the nonzero coefficients $m_0, m_1, \ldots, m_k$ to be smaller than β.*

Proof. We did not use any special properties of ω in the proof of Theorem 3.31, so we only need to modify Lemma 3.30. If we slightly change its claim, only for decreasing exponents $\gamma_0 > \gamma_1 > \cdots > \gamma_k$ and coefficients $m_i < \beta$, we can prove it by transfinite induction on γ_0. □

Remark. If we restrict ourselves only to natural numbers, we obtain the familiar theorem about expanding natural numbers using powers of a base $b > 1$.

3.3 Countable and Uncountable Ordinals

We saw earlier that for all natural numbers n, it holds that $n+\omega = n\cdot\omega = n^\omega = \omega$. We also proved that the functions corresponding to these basic operations,

$$A_n(\xi) = n + \xi, \qquad M_n(\xi) = n \cdot \xi, \qquad E_n(\xi) = n^\xi,$$

are normal. Theorem 3.14 claims that each of them has infinitely many fixed points. It is easy to see that no (nonzero) natural number is a fixed point, and above we have observed that ω is a fixed point of all of them. It is, in fact, the smallest (nonzero) fixed point. Notice that this makes intuitive sense. When restricted to natural numbers, these are all fast growing functions ($A \ll M \ll E$), so we need a new concept (countable infinity) to find a fixed point.

Now consider what would happen if we replaced n with ω and tried to find a (nonzero) fixed point of these new ω-functions. Theorem 3.14 claims that the smallest such fixed points are:

- $F_A = \sup\{0, \omega, \omega+\omega, \omega+\omega+\omega, \omega\cdot 4, \omega\cdot 5, \dots\} = \omega\cdot\omega$,
- $F_M = \sup\left\{1, \omega, \omega\cdot\omega, \omega\cdot\omega\cdot\omega, \omega^4, \omega^5, \dots\right\} = \omega^\omega$,
- $F_E = \sup\left\{1, \omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots\right\}$, and we no longer have notation to describe this number; we will denote it as ε_0.

The question is: did we escape the countable infinity represented by ω? No, we will soon see that all of these numbers are, in fact, still countable. Nonetheless, we have stumbled upon something important. The last number, $\varepsilon_0 = \omega^{\varepsilon_0}$, is closely connected to Peano arithmetic, and we will also use it when proving Goodstein's theorem. It also gives rise to an entire class of ordinals called the *epsilon numbers*.

3.3.1 Epsilon Numbers

Definition 3.34. An ordinal ξ is an *epsilon number* if it is a fixed point of the normal function $\xi \mapsto \omega^\xi$. That is, if $\xi = \omega^\xi$. Theorem 3.14 (iv) asserts the existence of a bijective normal function $\varepsilon : \mathrm{On} \to \{\xi \,|\, \xi = \omega^\xi\}$ enumerating the epsilon numbers. We denote by ε_β the ordinal $\varepsilon(\beta)$.

Proposition 3.35. *For any ordinal β, it holds that*

(i) $\varepsilon_0 = \sup\left\{1, \omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots\right\}$,

(ii) $\varepsilon_{\beta+1} = \sup\left\{1, \varepsilon_\beta, {\varepsilon_\beta}^{\varepsilon_\beta}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}}, \dots\right\}$,

(iii) $\varepsilon_\beta = \sup\{\varepsilon_\delta \,|\, \delta < \beta\}$, *whenever β is a limit ordinal.*

Proof. We prove the theorem by transfinite induction on β. (i) This is the definition of ε_0. (iii) Holds because the epsilon function is normal. (ii) Following Theorem 3.14, we know that $\varepsilon_{\beta+1}$ is the limit of the sequence

$$\varepsilon_\beta + 1, \omega^{\varepsilon_\beta+1}, \omega^{\omega^{\varepsilon_\beta+1}}, \omega^{\omega^{\omega^{\varepsilon_\beta+1}}}, \dots$$

Let α_n denote the element with index $n \in \omega$. Define a different sequence for $n \geq 2$ as $\alpha'_2 := \varepsilon_\beta^{\ \omega}$ and $\alpha'_{n+1} := \varepsilon_\beta^{\alpha'_n}$. Clearly,

$$\sup\{\alpha'_n \,|\, n \geq 2\} = \sup\left\{\varepsilon_\beta^{\ \omega}, \varepsilon_\beta^{\varepsilon_\beta^{\ \omega}}, \varepsilon_\beta^{\varepsilon_\beta^{\varepsilon_\beta^{\ \omega}}}, \dots\right\} = \sup\left\{1, \varepsilon_\beta, \varepsilon_\beta^{\varepsilon_\beta}, \varepsilon_\beta^{\varepsilon_\beta^{\varepsilon_\beta}}, \dots\right\}.$$

We will use induction on n to show $\alpha_n = \alpha'_n$ for all $n \geq 2$:

$$\begin{aligned}
\alpha_1 &= \omega^{\varepsilon_\beta+1} = \omega^{\varepsilon_\beta} \cdot \omega = \varepsilon_\beta \cdot \omega \\
\alpha_2 &= \omega^{\omega^{\varepsilon_\beta+1}} = \omega^{(\varepsilon_\beta \cdot \omega)} = (\omega^{\varepsilon_\beta})^\omega = \varepsilon_\beta^{\ \omega} = \alpha'_2 \\
\alpha_3 &= \omega^{\omega^{\omega^{\varepsilon_\beta+1}}} = \omega^{\varepsilon_\beta^{\ \omega}} = \omega^{\varepsilon_\beta^{\ 1+\omega}} = \omega^{\varepsilon_\beta \cdot \varepsilon_\beta^{\ \omega}} = (\omega^{\varepsilon_\beta})^{\varepsilon_\beta^{\ \omega}} = \varepsilon_\beta^{\varepsilon_\beta^{\ \omega}} = \alpha'_3 \\
\alpha_{n+2} &= \omega^{\alpha_{n+1}} = \omega^{\alpha'_{n+1}} = \omega^{\varepsilon_\beta^{\alpha'_n}} = \omega^{\varepsilon_\beta^{1+\alpha'_n}} = \omega^{\varepsilon_\beta \cdot \varepsilon_\beta^{\alpha'_n}} = (\omega^{\varepsilon_\beta})^{\alpha'_{n+1}} = \varepsilon_\beta^{\ \alpha'_{n+1}} = \alpha'_{n+2}
\end{aligned}$$

Hence $\varepsilon_\beta = \sup\{\alpha_n \,|\, n < \omega\} = \sup\{\alpha'_n \,|\, 2 \leq n < \omega\}$. □

Lemma 3.36 (AC_ω)**.** *A countable union of countable sets is countable. Specifically, if β and γ_α for $\alpha < \beta$ are countable ordinals, then $\gamma = \sup\{\gamma_\alpha \,|\, \alpha < \beta\}$ is also a countable ordinal.*

Remark. AC_ω denotes the *axiom of countable choice*, which states that every countable set has a choice function.

Proof. Let $A = \langle A_n \,|\, n \in I \rangle$ be a countable collection of sets, WLOG $I = \omega$, such that all A_n are countable, and denote $S := \bigcup A$. We will define an injection $g : S \to \omega \times \omega$ (here, $\omega \times \omega$ is countable[6]). Since each A_n is countable, it injects into ω, and we can *choose* an injection $j_n : A_n \to \omega$ (and because there are only countably many sets A_n, we are making only countably many choices). For an element $a \in S$, define

$$n_a := \min\{n \in \omega \,|\, a \in A_n\}.$$

This number indicates in which A_n does a first appear in. Notice that more elements $a \in S$ can have the same number n_a, but that $j_{n_a}(a)$ uniquely identifies a among these elements (since j_{n_a} is injective). Hence, we can define an injection $g : a \mapsto (n_a, j_{n_a}(a))$. □

From now on, we will generally assume AC_ω.

Lemma 3.37 (AC_ω)**.** *The ordinal ω^α is countable $\iff$ α is countable.*

Proof. We first prove '$\Leftarrow$' by transfinite induction on α. If $\alpha = 0$, it holds. Now suppose the claim holds for a countable α and consider the ordinal $\omega^{\alpha+1} = \omega^\alpha \cdot \omega$. This is countable because it is the order type of the set $\omega \times \omega^\alpha$, which is countable since it is the cartesian product of two countable sets. Finally, if α is a countable limit ordinal, then

$$\omega^\alpha = \sup\{\omega^\delta \,|\, \delta < \alpha\} = \bigcup\{\omega^\delta \,|\, \delta < \alpha\}.$$

Because α is countable and all ω^δ are countable ordinals (induction hypothesis), Lemma 3.36 implies that ω^α is countable as well.

[6] Prove that the Cartesian product of finitely many countable sets is countable. Hint: prime numbers might help. Then try proving it without the use of primes.

We prove '$\Rightarrow$' by contraposition. Suppose that α is uncountable; since $\xi \mapsto \omega^\xi$ is a normal function, it grows at least as fast as the identity function, and so $\omega^\alpha \geq \alpha$. We can now easily define an injection $\alpha \to \omega^\alpha$, showing that ω^α is uncountable. $\square$

Lemma 3.38. *The epsilon number ε_0 is countable.*

Proof. By definition, ε_0 is the limit of the sequence $\alpha_0 = 1$ and $\alpha_{n+1} = \omega^{\alpha_n}$. By induction on n and using the previous lemma, one can show that all α_n are, in fact, countable. This implies that ε_0 is a countable union of countable ordinals and is therefore countable as well. $\square$

Note on ordinal notations We have proven the previous lemma via a statement that requires AC_ω, because it made our job easier (and will continue to do so greatly down the line). However, we could have shown that ε_0 is countable by realizing that every $\alpha < \varepsilon_0$ has a finite *hereditary* Cantor normal form; therefore, we can use prime numbers to encode α as a unique natural number, hence constructing an injection $\varepsilon_0 \to \omega$. Hereditary CNF simply means that if any of the powers γ_i are ordinals larger than ω, then we express them in CNF as well, and we repeat the process inductively. For example:

$$\alpha = \omega^{\omega^{\omega+1}+\omega^2\cdot 3+5} + \omega^{\omega\cdot 2+1} + \omega \cdot 2 + 7.$$

But ε_0 cannot be represented by a finite hereditary CNF, since $\varepsilon_0 = \omega^{\varepsilon_0}$ is its CNF, which is self-referential.

This concept can be generalized: if λ is a large ordinal, and every $\alpha < \lambda$ can be represented as a finite syntactic structure over some finite alphabet, then we can encode α as a unique natural number, constructing an injection $\lambda \to \omega$. This concept is called *ordinal notations*, and it does not require any kind of choice. However, the larger the ordinal λ, the more complicated the finite structure and encoding become. It is crucial to understand that when we use AC_ω to prove the countability of a large ordinal (like ε_0 or Γ_0 later), we are using it to simplify the process, not because we *have to* use it.

However, general statements like Lemma 3.37 or the following theorem do require the use of countable choice and cannot be proved in bare ZF.

Theorem 3.39 (AC_ω)**.** *The epsilon number ε_β is countable $\iff$ β is countable.*

Proof. We first prove '$\Leftarrow$' by transfinite induction on β. The base case $\beta = 0$ has been verified by the previous lemma. Suppose the claim holds for a countable ordinal β; that is, ε_β is countable, and we want to show that $\varepsilon_{\beta+1}$ is countable as well. Theorem 3.14 claims that $\varepsilon_{\beta+1}$ is the limit of the sequence $\alpha_0 = \varepsilon_\beta + 1$ and $\alpha_{n+1} = \omega^{\alpha_n}$. Note that α_0 is countable since ε_β is countable. One can now use induction on n and Lemma 3.37 to show that all α_n are in fact countable. This implies that $\varepsilon_{\beta+1}$ is a countable union of countable ordinals and is therefore also countable.

Finally, assume that β is a countable limit ordinal. Because the epsilon function $\beta \mapsto \varepsilon_\beta$ is normal, it holds that

$$\varepsilon_\beta = \sup\{\varepsilon_\delta \mid \delta < \beta\} = \bigcup\{\varepsilon_\delta \mid \delta < \beta\}.$$

Since β is countable, this is a countable union of countable ordinals (induction hypothesis), so it is countable as well.

One can prove '⇒' in the same manner as in Lemma 3.37. □

The theorem we have just proven places us in a difficult position. Does an uncountable ordinal even exist? If we assume the axiom of choice, then it is fairly easy to find one: just well-order the uncountable set $\mathcal{P}(\omega)$ and take its order type. Finding one in $\mathsf{ZF} + \mathsf{AC}_\omega$ seems to be much more difficult.

3.3.2 Veblen Hierarchy

We know that combining countable ordinals using the standard operations defined above produces more countable ordinals. The best tool for constructing large ordinals we currently have are the epsilon numbers (and, moreover, Theorem 3.14 implies that for any ordinal, there is a larger epsilon number), but it seems like they will not help us either. Consider the sequence

$$\gamma_0 = \varepsilon_0,\ \gamma_{n+1} = \varepsilon_{\gamma_n} \qquad \longrightarrow \qquad \varepsilon_0, \varepsilon_{\varepsilon_0}, \varepsilon_{\varepsilon_{\varepsilon_0}}, \varepsilon_{\varepsilon_{\varepsilon_{\varepsilon_0}}}, \ldots$$

The largest number we can currently construct is the limit of this sequence; the first fixed point of the epsilon function $\xi \mapsto \varepsilon_\xi$, a number denoted as ζ_0. However, this number, while enormously large, is still countable. The reason is that all the terms γ_n are countable (by induction and the previous theorem); hence, ζ_0 is a countable union of countable ordinals and is thus also countable.

We could define *zeta* (ζ) *numbers* in a similar fashion to how we defined epsilon numbers; however, for the same reasons that epsilon numbers with countable indices are countable, we would arrive at the conclusion that any zeta number with a countable index is still countable.

We could even create an entire hierarchy of these special fixed-point numbers. The bottom tier would be $\varphi_0(\beta) = \omega^\beta$; the second tier would be the epsilon numbers $\varphi_1(\beta) = \varepsilon_\beta$; the third tier would be the zeta numbers $\varphi_2(\beta) = \zeta_\beta$; the next one would be the so-called *eta numbers* $\varphi_3(\beta) = \eta_\beta$, and so on. The tiers are defined in such a way that the values of φ_{n+1} are the fixed points of φ_n. We could now consider the ordinal

$$\Lambda \coloneqq \sup\{\varphi_n(0) \mid n < \omega\}.$$

However, this ordinal is *still countable*, as it is a countable union of countable ordinals.

The functions φ_n we have just constructed are called the *Veblen functions*, and they can be generalized for arbitrary ordinal indices.

Definition 3.40 (Veblen hierarchy, 1908)**.** The functions $\varphi_\alpha : \mathrm{On} \to \mathrm{On}$ are defined for all ordinals $\alpha \geq 0$ recursively as:

(i) $\varphi_0(\beta) = \omega^\beta$,

(ii) $\varphi_{\alpha+1}(\beta)$ is the βth fixed point of φ_α, starting at $\beta = 0$.

(iii) when α is a limit ordinal, we define $\varphi_\alpha(\beta)$ as the βth simultaneous fixed point of all the φ_δ for $\delta < \alpha$, also starting at $\beta = 0$.

Observation 3.41. *The function φ_α is normal for every α.*

Proof. By transfinite induction on α. It holds for $\alpha = 0$ because exponentiation is normal in the second argument. If $\alpha = \gamma + 1$ is isolated, then φ_α enumerates the fixed points of the normal function φ_γ and is by Theorem 3.14 (iv) normal. If α is isolated, it enumerates the simultaneous fixed points of the normal functions φ_δ for $\delta < \alpha$ and so φ_α is normal by Theorem 3.15 (iv). □

Exercise 9. Show that if $\gamma > \alpha$, then $\varphi_\alpha(\varphi_\gamma(\beta)) = \varphi_\gamma(\beta)$ for any β.

This demonstrates that the values of $\varphi_{\alpha+1}$ are not only fixed points of φ_α, but they are fixed points of all φ_δ for $\delta \leq \alpha$. This means that we could have used condition (iii) to define φ_α for all ordinals α, not only for limit ordinals.

Exercise 10. Show that the ordinal function $\alpha \mapsto \varphi_\alpha(0)$ is normal.

Definition 3.42. An ordinal worth noting is the *Feferman–Schütte ordinal* Γ_0, defined as the first fixed point of the function $\alpha \mapsto \varphi_\alpha(0)$, or equivalently, as the limit of the sequence $\gamma_0 = \varphi_0(0)$, $\gamma_{n+1} = \varphi_{\gamma_n}(0)$, that is

$$\gamma_0 = 1, \gamma_1 = \varphi_1(0) = \varepsilon_0, \gamma_2 = \varphi_{\varepsilon_0}(0) = \varphi_{\varphi_{\varphi_0(0)}(0)}(0), \gamma_3 = \varphi_{\varphi_{\varphi_{\varphi_0(0)}(0)}(0)}(0), \ldots$$

It is one of the most famous ordinals in logic, and we will attempt to provide an explanation of why in Section 3.4.5.

Exercise 11. Show that Γ_0 is the first ordinal $\gamma > 0$ closed under φ. That is, the least $\gamma > 0$ such that for all $\alpha, \beta < \gamma$ we have $\varphi_\alpha(\beta) < \gamma$.

In other words, Γ_0 is the first ordinal that cannot be reached from below via repeated application of the Veblen functions. Does that mean that we have finally found an uncountable ordinal?

Exercise 12. Show that for all ordinals $\alpha < \Gamma_0$ it holds that $\varphi_\alpha(\Gamma_0) = \Gamma_0$.

That is, all φ_α attain the value Γ_0 at the same time! Even the very "slow" growing $\varphi_0(\beta) = \omega^\beta$ catches up to functions like $\varphi_{\varepsilon_0}(\beta)$ or $\varphi_{\varphi_{\varepsilon_0}(0)}(\beta)$, and they all momentarily synchronize at $\beta = \Gamma_0$.

Theorem 3.43 (AC_ω)**.** *The ordinal $\varphi_\alpha(\beta)$ is countable $\iff$ α, β are countable.*

Proof. For the '$\Leftarrow$' direction, we use transfinite induction on α. Notice that Lemma 3.37 is our base case (when $\alpha = 0$). If $\alpha = \gamma + 1$ is isolated, the proof is essentially the same as the proof of Theorem 3.39. If α is a countable limit ordinal, then we prove the claim by transfinite induction on β.

- If $\beta = 0$, then $\varphi_\alpha(0)$ is by Theorem 3.15 (ii) equal to the limit of the sequence

$$\gamma_0 = 0, \qquad \gamma_{n+1} = \sup\{\varphi_\delta(\gamma_n) \mid \delta < \alpha\}.$$

  Using our outer induction hypothesis (for α), one can show by induction on n that all γ_n are countable. Thus $\varphi_\alpha(0)$ is a countable union of countable ordinals and is therefore also countable.

- If $\beta = \beta' + 1$ is isolated, then the proof is almost identical to the case when $\beta = 0$. One just starts with $\gamma_0 = \varphi_\alpha(\beta') + 1$.

- If β is a limit ordinal, then since φ_α is normal we have that $\varphi_\alpha(\beta) = \sup\{\varphi_\alpha(\delta) \mid \delta < \beta\}$. Because we assume that β is countable, $\varphi_\alpha(\beta)$ is thus a countable union of countable ordinals (inner induction hypothesis) and is countable as well.

For the '$\Rightarrow$' direction, we need to show that $\varphi_\alpha(\beta) \geq \alpha, \beta$; so if either α or β is uncountable, then $\varphi_\alpha(\beta)$ is also uncountable. Clearly $\varphi_\alpha(\beta) \geq \beta$ because φ_α is a normal function. Exercise 10 implies that $\varphi_\alpha(0) \geq \alpha$, and from here we have that $\varphi_\alpha(\beta) \geq \varphi_\alpha(0) \geq \alpha$. □

Corollary 3.44. *The Feferman–Schütte ordinal* Γ_0 *is countable, as it is a countable union of countable ordinals.*

Remark. As mentioned in Section 3.3.1, the axiom of countable choice is not required to show that Γ_0 is countable, but it makes the task easier. If we wanted to prove it in ZF, we would show that the Veblen functions $\varphi_\alpha(\beta)$ provide a way of expressing every $\gamma < \Gamma_0$ as a unique finite syntactic structure. For more details about the Veblen hierarchy, see Sections 6.5, 7, and 8 of [9].

What we have shown in this section demonstrates an important concept in set theory when assuming AC_ω: one cannot reach uncountable infinity by starting from ω and applying ordinal operations such as addition, multiplication, exponentiation, finding fixed points of normal functions, and taking suprema — all the while utilizing only the ordinals we have already constructed along the way.

3.3.3 Hartogs' Theorem

Does that mean that all hope is lost and there are no uncountable ordinals? Thankfully, no. The following theorem gives us a way out.

Theorem 3.45 (Hartogs, 1915)**.** *For any set x, there exists an ordinal η such that there is no injection $\eta \to x$. The least such η is called the* Hartogs number *of x.*

Proof (cf. [8]*).* Consider the set (why is this a set?)

$$\mathcal{W} = \{(A, <_R) \mid A \subseteq x \text{ and } <_R \text{ is a well-ordering of } A\}.$$

We can use replacement to construct the set

$$S = \{\alpha \in \mathrm{On} \mid \text{there exists } (A, <_R) \in \mathcal{W} \text{ order-isomorphic to } \alpha\}$$

by assigning to each $(A, <_R)$ its order type.

But this set is exactly the Hartogs number of x. Notice that S is transitive: if $\alpha \in S$ and $\gamma < \alpha$, then $\alpha \in S$ as well. A transitive set of ordinals is again an ordinal (Lemma 2.9), so S is an ordinal number η. Furthermore, there is no injection from η into x, because if there were, then we would get the contradiction that $\eta \in \eta$. And finally, η is the least such ordinal. If $\alpha < \eta$, then also $\alpha \in \eta$, and there is an injection $\alpha \to x$. □

Remark. It is crucial to note that the theorem we just proved, which gives us the Hartogs number as a von Neumann ordinal, is more powerful than Hartogs' original 1915 result. Hartogs, working in Z (proposed in 1908 by Zermelo, containing the axioms of ZF except replacement and foundation), only proved the existence of *a* well-ordered set that could not be injected into x; but he did not—and could not—show it was a von Neumann ordinal. The general theorem that "every well-ordered set is isomorphic to a unique von Neumann ordinal" is itself not provable in Z and requires replacement (see proof of Theorem 2.13). We will now use this modern, replacement-based construction to construct an uncountable ordinal as the Hartogs number of ω. It is this very step, guaranteeing that the collection of all countable ordinals is a set, that fails in Z and was one of the motivations for Fraenkel and Skolem to propose the axiom of replacement in 1922.

Notice that the theorem does not say that $x \prec \eta$, because this does not in general hold without AC. However, if x is well-ordered, then it has an order type α, and we can compare α with η.

This allows us to access an uncountable ordinal. Let ω_1 be the Hartogs number of ω. That is, the first ordinal with the property $\omega_1 \not\preceq \omega$, or in other words, the *first uncountable ordinal*,[7] and we can write

$$\omega_1 = \{\alpha \in \mathrm{On} \mid \alpha \preceq \omega\}$$

Exercise 13. Show that an ordinal α is uncountable (that is $\alpha \not\preceq \omega$) $\iff$ $\omega \prec \alpha$. Note that this is not true for general sets in ZF, we would need to accept AC_ω.

It is almost impossible to grasp just how unfathomably large ω_1 is. The entire vast, complex, mind-boggling hierarchy of ordinals described by the Veblen functions up to Γ_0 is still just a tiny, countable speck at the absolute "bottom" of the ordinal line from the perspective of ω_1.

Exercise 14. Show that for any countable ordinal α, it holds that $\varphi_\alpha(\omega_1) = \omega_1$, specifically $\omega^{\omega_1} = \omega_1$ and $\varepsilon_{\omega_1} = \omega_1$. Furthermore show that $\varphi_{\omega_1}(0) = \omega_1$.

Realize that there was nothing special about the choice of ω. We can apply the same process to ω_1 to get ω_2, and continue doing this to construct larger and larger ordinals (in the sense of cardinality).

Definition 3.46. For an ordinal α we define ω_α recursively as

(i) $\omega_0 := \omega$,

(ii) $\omega_{\alpha+1}$ is the Hartogs number of ω_α,

(iii) $\omega_\lambda := \sup\{\omega_\alpha \mid \alpha < \lambda\}$ for limit ordinals λ.

Observation 3.47. *The number ω_α is the first ordinal that is larger (in the sense of cardinality) than all previous ω-numbers.*

This definition foreshadows the section about cardinal numbers, where we will encounter these omega numbers again and explore their properties in depth.

Hartogs numbers also allow us to finally prove that the trichotomy principle implies AC. In fact, this was the original motivation behind Hartogs' theorem.

[7] The ordinal ω_1 is also commonly denoted as Ω.

Theorem 3.48. *The trichotomy principle implies the well-ordering principle.*

Proof. Let x be an arbitrary set, and let η be its Hartogs number. Apply the trichotomy principle to x and η. One of the following holds:

(a) $x \preceq \eta$, there is an injection $x \to \eta$, or

(b) $\eta \preceq x$, there is an injection $\eta \to x$.

The second case is impossible due to the defining property of η. Hence, there exists an injection $f : x \to \eta$. We can now well-order x by inheriting the order of η by f. $\square$

3.4 Peano Arithmetic

To understand this section, the reader should be familiar with the basic notions of logic, including concepts such as language, theory, model, etc. Explanations of these concepts can be found in the lecture notes [3] for the course NAIL062.

3.4.1 Peano Axioms

Peano Arithmetic, denoted PA, is the standard axiomatic theory of the natural numbers. In ZFC, we have encountered the set of natural numbers, ω, constructed as the set of finite von Neumann ordinals. This is no coincidence; the set ω, together with the restrictions of operations of ordinal arithmetic to ω, serves as the *standard model* for PA, denoted by $\mathcal{N}$.

Our study of ordinal arithmetic in Section 3.2 has already established that these operations, when restricted to finite ordinals, are commutative and satisfy all the familiar properties of elementary arithmetic. The axioms of PA can therefore be seen as a precise, first-order logic attempt to capture the properties of this standard model.

Definition 3.49 (PA, [3])**.** The language of PA is $\mathcal{L}_{PA} = \langle 0, S, +\,, \cdot\,, \leq \rangle$ with equality. The base axioms of PA are the following formulas:

$$\begin{aligned} &\neg Sx = 0 && x \cdot 0 = 0 \\ &Sx = Sy \implies x = y && x \cdot Sy = x \cdot y + x \\ &x + 0 = x && \neg x = 0 \implies (\exists y)(x = Sy) \\ &x + Sy = S(x + y) && x \leq y \iff (\exists z)(z + x = y) \end{aligned}$$

These axioms alone yield the much weaker *Robinson Arithmetic* (Q). It cannot prove, for example, the commutativity or associativity of addition or multiplication, or the transitivity of order. To obtain PA, we need to add the *Axiom Schema of Induction.* That is, for each $\mathcal{L}_{PA}$-formula $\varphi(x, \vec{y})$, the following axiom is added:

$$\big(\varphi(0, \vec{y}) \wedge (\forall x)(\varphi(x, \vec{y}) \Rightarrow \varphi(Sx, \vec{y}))\big) \implies (\forall x)\varphi(x, \vec{y}) \tag{3.3}$$

Remark. The last axiom schema should seem similar to the induction principle on ω from set theory:

$$(\forall X \subseteq \omega)\Big(\big(0 \in X \wedge (\forall x)(x \in X \Rightarrow x \cup \{x\} \in X)\big) \implies X = \omega\Big).$$

However, the axiom schema of induction is a weaker version, as it is a first-order logic attempt to simulate a second-order logic axiom with an axiom schema. The familiar induction principle could be expressed with the following second-order $\mathcal{L}_{PA}$-formula

$$(\forall X)\Big(\big(X(0) \wedge (\forall x)(X(x) \Rightarrow X(Sx))\big) \implies X = (\forall x)X(x)\Big).$$

By adding it to PA, we would obtain the much stronger second-order theory PA_2.

Here X represents (any) unary relation; that is, a subset of the universe. The important distinction is that (3.3) provides an infinite collection of axioms, one for each subset of the universe that is *definable* by a $\mathcal{L}_{PA}$-formula φ.

This restriction is the source of PA's most profound properties and limitations. For example, PA_2 is categorical; that is, it has only one model (up to isomorphism) — the standard model $\mathcal{N}$. On the other hand, PA allows the existence of other non-standard models.

3.4.2 Models of Arithmetic

We have already mentioned that the *standard model* of PA is the $\mathcal{L}_{PA}$-structure $\mathcal{N} = (\omega, 0^{\mathcal{N}}, S^{\mathcal{N}}, +^{\mathcal{N}}, \cdot^{\mathcal{N}}, \leq^{\mathcal{N}})$, where the domain is the set ω, the interpretation of the symbol '0' is $0^{\mathcal{N}} = \varnothing$, the successor of x is $S^{\mathcal{N}}(x) = x \cup \{x\}$, and $+^{\mathcal{N}}$, $\cdot^{\mathcal{N}}$ and $\leq^{\mathcal{N}}$ are the operations of ordinal arithmetic restricted to ω.

Theorem 3.50. *There exist countable models of* PA *that are not isomorphic to* $\mathcal{N}$.

Proof sketch. By the Compactness Theorem. We extend $\mathcal{L}_{PA}$ with a new constant symbol c. Consider the theory $T = \mathsf{PA} \cup \{c > \bar{n} \mid n \in \omega\}$, where $\bar{n}$ is the $\mathcal{L}_{PA}$-term $S(S(\ldots S(0)\ldots))$ (n times). Any finite subset $T_0 \subset T$ is satisfiable: we take $\mathcal{N}$ as the model and interpret c as a standard natural number larger than any numeral $\bar{n}$ explicitly mentioned in T_0. By the Compactness Theorem, T has a model $\mathcal{M}$. This $\mathcal{M}$ must be a model of PA, but the interpretation of c is a "non-standard" number, an element larger than all standard elements $S^n(0)$. Thus, $\mathcal{M} \not\cong \mathcal{N}$. □

All countable non-standard models $\mathcal{M}$ share a common structure: they begin with an initial segment isomorphic to ω (the standard part), which is then followed by a collection of "blocks" of non-standard numbers. This "pathology" of PA is not merely set-theoretic, but also computational.

Theorem 3.51 (Tennenbaum, 1959)**.** *No countable non-standard model of* PA *is recursive.*

This implies that in any non-standard model $\mathcal{M}$, the operations $\oplus$ and $\otimes$ (the interpretations of $+$ and $\cdot$) are not computable functions. Even if the domain of $\mathcal{M}$ is ω, the operations themselves cannot be implemented by an algorithm. The induction schema, while syntactically "weaker" than its second-order counterpart, thus imposes enormous computational complexity on any "non-standard" structure that satisfies it, effectively isolating the standard model as the only computationally tractable one.

3.4.3 Gödel's Incompleteness Theorems

When working with a formal theory, it is natural to ask what statements we can prove from its axioms. If a theory T can prove the sentence ψ, we write $T \vdash \psi$. A theory is *consistent* if it is free from contradictions, meaning it is impossible to prove both a statement ψ and its negation $\neg\psi$ from its axioms; or equivalently, if it has a model. A consistent theory is *complete* if it has an "opinion" on every statement, meaning for every sentence ψ in its language, the theory can prove either ψ or $\neg\psi$. If it cannot do either, it is said to be *incomplete*, and ψ is said to be *independent* in T. Equivalently, ψ is independent in T, if it holds in some models of the theory but does not hold in others.

Probably the most influential result linking these concepts together with PA are the famous Incompleteness Theorems, published by Kurt Gödel [8] [11] in 1931. Veritasium has an amazing video [25] that provides an intuitive explanation of this topic. We provide only a simplified explanation of these profound results; for more details and proofs, refer to [3].

Despite its limitations, PA is a remarkably powerful theory. Its expressive power is sufficient to represent all computable (recursive) functions. This strength is the key to PA's own undoing. It allows for the *arithmetization of syntax* (Gödel numbering), whereby the syntax of $\mathcal{L}_{PA}$ (terms, formulas, proofs) can be uniquely encoded as natural numbers. Syntactic operations (like substitution) and relations (like "is a proof of") become recursive functions and relations on these numbers. Crucially, this allows for the creation of a provability predicate.

Definition 3.52. There exists an $\mathcal{L}_{PA}$-formula $\mathrm{Prov}_{\mathsf{PA}}(x)$ such that for any sentence ϕ it holds that $(\mathsf{PA} \vdash \phi) \iff \mathrm{Prov}_{\mathsf{PA}}(\ulcorner\phi\urcorner)$. Here, $\ulcorner\phi\urcorner$ denotes the Gödel number of ϕ. The formula $\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\phi\urcorner)$ is: "there exists x such that x is the Gödel number of a proof of the sentence with Gödel number $\ulcorner\phi\urcorner$."

This predicate allows the theory to "talk about" its own provability, leading directly to sentences that self-reference and assert their own unprovability.

Theorem 3.53 (Gödel's First Incompleteness Theorem, 1931)**.** *If* PA *is consistent, then it is incomplete.*

Proof sketch. Consider a sentence $\mathfrak{g}$ (the *Gödel sentence*) saying: "there is no x such that x is the Gödel number of a proof of the sentence with Gödel number $\ulcorner\mathfrak{g}\urcorner$." Notice that $(\mathsf{PA} \vdash \mathfrak{g}) \iff \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)$. Hence if $\mathsf{PA} \vdash \mathfrak{g}$, then PA is inconsistent. Therefore, if PA is consistent, then $\mathsf{PA} \nvdash \mathfrak{g}$, and it is incomplete. □

As a corollary of this theorem, Gödel achieved his second result.

Theorem 3.54 (Gödel's Second Incompleteness Theorem, 1931)**.** PA *cannot prove its own consistency.*

[8] The life of Kurt Gödel (1906–1978) is a fascinating story. Born in Brno, he left for Vienna at the age of eighteen to study mathematics and logic. At twenty-four, he proved his incompleteness theorems, which formed the basis of his doctoral dissertation. He later emigrated to the United States following the rise of Nazism. Albert Einstein regarded Gödel as the greatest logician since Aristotle and once remarked that the only reason he went to his office was to have the privilege of walking home with Gödel. Yet Gödel's life was not without darkness: he struggled with psychological illness throughout adulthood and ultimately died of self-starvation, driven by the paranoid belief that someone was trying to poison him. Perhaps the most detailed account of Gödel's life (as of the writing of this text) can be found in [2].

Proof sketch. Let Con(PA) be the $\mathcal{L}_{PA}$-sentence $\neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\perp\urcorner)$ (where $\perp$ is a contradiction, e.g., $0 = S0$). That is, Con(PA) is true if and only if PA is consistent. Because Gödel formalized the entire proof of the previous theorem in PA (using Gödel numbers), his first theorem can be expressed as

$$\mathsf{PA} \vdash \big(\mathrm{Con}(\mathsf{PA}) \implies \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)\big).$$

This together with the equivalence $(\mathsf{PA} \vdash \mathfrak{g}) \iff \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)$ gives

$$\mathsf{PA} \vdash \big(\mathrm{Con}(\mathsf{PA}) \implies \mathfrak{g}\big).$$

Now, suppose for contradiction that PA could prove its own consistency. Combining this with the last formula gives $\mathsf{PA} \vdash \mathfrak{g}$, but this is a contradiction, since (the end of the previous proof) if PA is consistent, then $\mathsf{PA} \nvdash \mathfrak{g}$. □

Gödel's original formulation of these theorems did not, in fact, talk about PA, but about a system he called P, a close relative of PA. Gödel then had to make a philosophical assumption. He argued that any other system "related" to, and at least as strong as P (and therefore capable of arithmetic), would also be capable of producing a Gödel sentence $\mathfrak{g}$; thus, his incompleteness theorems would apply to this system as well. This was a strong, intuitive argument, but he could not formally prove it.

The missing piece was provided in 1936 by Alan Turing [24], who formalized the notion of computability using the Turing machine, which made a formal proof of Gödel's conjecture possible.

Theorem 3.55 (Generalized Gödel's Incompleteness Theorems)**.** *For any consistent, recursively axiomatized theory T, it holds that:*

(1) *If T is an extension of Robinson arithmetic* Q*, then T is incomplete.*

(2) *If T is an extension of Peano arithmetic* PA*, then T cannot prove its own consistency.*

Remark. Recursively axiomatized means that there is an algorithm (Turing machine) that, for every input formula φ, halts and answers whether φ is an axiom of T. The condition that T is an extension of Q (or PA) essentially means that T is at least as powerful as Q (or PA). For example, PA is an extension of Q.

Corollary 3.56. *It is impossible to prove the consistency of* ZFC *inside* ZFC*.*

An example of an independent statement in ZFC is the famous continuum hypothesis CH, claiming that there is no set x such that $\omega \prec x \prec \mathcal{P}(\omega)$. Similarly, AC can be shown to be independent in ZF, meaning that if ZF is consistent, then ZFC is as well.

In 1940, Gödel showed that neither AC can be disproved from ZF, nor CH from ZFC, by constructing the *constructible universe*, a model of ZF in which both AC and CH hold. This model begins with the empty set and adds only those sets that are definable from previous ones, thus forming the minimal universe compatible with the axioms. Later, in 1963, Paul Cohen showed that CH cannot be proved from ZFC by developing the method of *forcing*, which allowed him to construct a model of ZFC in which CH fails. Through a different forcing argument, he likewise obtained a model of ZF that violates AC.

3.4.4 Consistency and the Connection with ε_0

Gödel's second theorem seems to place us in a difficult position: a consistency proof for PA must employ principles that transcend PA itself. While ZFC is far stronger than PA and easily proves $\mathrm{Con}(\mathsf{PA})$ (by exhibiting the model $\mathcal{N}$), this isn't a very "unilluminating" result. PA is a "finitary" theory, while ZFC is a wildly "infinitary" theory (it assumes the existence of various vast infinities). By proving the consistency of PA in ZFC, we base our proof on the assumption that ZFC is consistent. It would be better to find a weaker system that is still capable of proving $\mathrm{Con}(\mathsf{PA})$.

Theorem 3.57 (Gentzen's Consistency Proof,[9] 1936)**.** *The consistency of* PA *is provable in Primitive Recursive Arithmetic* PRA *(which by itself is weaker than* PA*), augmented with a schema for transfinite induction up to the ordinal* ε_0*.*

Gentzen's proof precisely identified the principle transcending PA that is required to prove its consistency. Recall from Section 3.3.1 that ε_0 is the first fixed point of the ordinal function $\alpha \mapsto \omega^\alpha$, the limit of the sequence $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$ Gentzen's result, $\mathsf{PRA} + \mathrm{TI}(\varepsilon_0) \vdash \mathrm{Con}(\mathsf{PA})$, thus establishes two facts:

(a) Proving the consistency of PA does not require the full power of ZFC; only transfinite induction up to a countable ordinal. That is, the assumption that ε_0 contains no infinite decreasing chains.

(b) The principle of transfinite induction up to ε_0, $\mathrm{TI}(\varepsilon_0)$, must be unprovable in PA (lest PA prove its own consistency).

Gentzen also showed that using any smaller ordinal $\alpha < \varepsilon_0$ is not enough. This calibrates the strength of PA with extraordinary precision. The collected strength of PA's infinite induction schema is exactly equivalent to the single principle of transfinite induction up to (but not including) ε_0. This is formalized in the concept of the *proof-theoretic ordinal*.

Theorem 3.58. *The* proof-theoretic ordinal *of* PA *is* $|\mathsf{PA}| = \varepsilon_0$*.*

This theorem has a twofold meaning that we can understand intuitively:

(a) What PA *can* prove: PA is strong enough to prove the well-foundedness of any recursive well-ordering $<_R$ on ω with order-type $\alpha < \varepsilon_0$.

(b) What PA *cannot* prove: PA is *not* strong enough to prove the well-foundedness of any recursive well-ordering $<_R$ on ω with order-type $\alpha \geq \varepsilon_0$.

Here, "recursive" means that there exists an algorithm that can answer whether $x <_R y$ or $y <_R x$ for all $x, y \in \omega$. Well-foundedness is the arithmetical statement that every nonempty subset of ω has a minimal element. Proving well-foundedness is thus equivalent to proving that transfinite induction "works" for that ordering (as there cannot be any infinite decreasing chains).

A point of confusion here might be the fact that any well-ordering is well-founded. But PA does not know that $<_R$ is a well-ordering; it only receives

[9] See [21] for a modern version of the proof. Moreover, [6] provides an alternative view on this result, and talks about the consistency of PA in general.

an "object," $<_R$, together with the instructions: "prove that the ordering you received is well-founded."

Therefore, PA can formalize proofs by transfinite induction up to any ordinal $\alpha < \varepsilon_0$, but it cannot justify the principle of transfinite induction up to ε_0 itself.

3.4.5 Limits of Predicative Mathematics

The discovery of set-theoretic paradoxes (such as Russell's and Burali-Forti's) in the early 20th century triggered the so-called *Grundlagenkrise*, or foundational crisis, in mathematics. The naive assumption that any property $\phi(x)$ could define a set $\{x \mid \phi(x)\}$ was shown to lead to contradictions. This prompted a range of philosophical responses, but most mathematicians eventually turned to the axiomatic framework of ZFC, which resolved the paradoxes by carefully restricting what counts as a set.

However, a significant objection came from mathematicians like Poincaré and, most notably, Hermann Weyl. They argued that the core problem was the use of *impredicative definitions*—definitions that define an object S by quantifying over a totality T that already includes S. This went directly against ZFC, as it uses impredicative definitions all the time. For example, the supremum of a set is defined as its least upper bound. In this case, the totality is the set of all upper bounds, and since the supremum is itself an upper bound, it is a member of that totality.

Weyl argued that such definitions were circular and potentially dangerous. He decided to rebuild mathematical analysis on a "safe" *predicative* basis, starting with his 1918 paper "The Continuum." He succeeded in developing a significant portion of classical analysis, but was unable to replicate everything. The big question became: "How much of mathematics can we *actually* recover using only predicative methods?" In the 1960s, Solomon Feferman and Kurt Schütte independently found the precise answer. We provide an intuitive interpretation of their result.

Theorem 3.59 (Feferman–Schütte, c. 1965)**.** *The proof-theoretic ordinal of "predicative mathematics" is the* Feferman–Schütte ordinal*, denoted* Γ_0*. It is the first fixed point of the function* $\alpha \mapsto \varphi_\alpha(0)$ *(see Section 3.3.2).*

This shows that Γ_0 is the first ordinal that cannot be proven to be well-founded by predicative means, just as ε_0 is the first ordinal that cannot be proven well-founded by the "finitistic" means of PA.

Remark. It should be noted that not all mathematicians agree on what "predicative mathematics" means exactly. It would be more accurate to say that Feferman and Schütte showed that Γ_0 is the first ordinal that cannot be proven well-founded by *certain* predicative means, and most people agree that those means are a reasonable interpretation of predicativity. This was later formalized in the 1970s by Friedman and Simpson in a formal system called ATR_0 (Arithmetical Transfinite Recursion). So what the theorem above really says is that the proof theoretic ordinal of ATR_0 is Γ_0.

The discipline of finding the proof-theoretic ordinals of theories is called *Ordinal Analysis*. For an introduction to the topic, I recommend [21]. Other notable sources are [7], [22] and [19].

3.5 Applications of Countable Ordinals

Gödel's independent sentences, $\mathfrak{g}$ and Con(PA), are meta-mathematical statements, not "natural" theorems of number theory. For decades, it was an open question whether any "ordinary" theorem of arithmetic or combinatorics was unprovable in PA, leading to speculation that Gödel's Incompleteness Theorem would not have meaningful implications to practical mathematics.

However, in 1977, Paris and Harrington [20] showed that a very natural variation of Ramsey's Theorem was true but not provable in PA. Five years later, in 1982, Kirby and Paris [18] showed that Goodstein's theorem, a statement purely about sequences of natural numbers, cannot be proven in PA either.

The second theorem presented in Kirby and Paris's 1982 paper establishes an analogous result, this time showing that a statement about the Hydra game is unprovable in PA.

3.5.1 Goodstein Sequences

Definition 3.60. The *hereditary base-n representation* of a natural number m is achieved by first writing m in base n, and then applying the procedure inductively to each exponent until there are no numbers larger than n.

Example. 100 in hereditary base-2 is $64+32+4 = 2^6+2^5+2^2 = 2^{2^2+2}+2^{2^2+1}+2^2$.

The Goodstein sequence of a natural number $m > 0$ is generated as follows.

- Start with $m_0 = m$.
- To get m_1, write m_0 in hereditary base-2 representation, replace all 2s with 3s, and then subtract 1 from the result.
- To get m_{n+1} from m_n, write m_n in hereditary base-$(n+2)$ representation, replace each occurrence of $(n+2)$ with $(n+3)$, and subtract 1.
- If ever $m_n = 0$, then $m_{n+1} = 0$.

For example, when we start with $m_0 = 3$, we get the sequence:

$$\begin{aligned}
m_0 &= 2^1 + 1 && = 3\\
m_1 &= 3^1 + 1 - 1 = 3^1 && = 3\\
m_2 &= 4^1 - 1 = 3 && = 3\\
m_3 &= 3 - 1 = 2 && = 2\\
m_4 &= 2 - 1 = 1 && = 1\\
m_5 &= 1 - 1 = 0 && = 0
\end{aligned}$$

Notice that at step $m_2 \to m_3$, the base used (4 at m_2) exceeded the value of $m_2 = 3$, which caused the sequence to start decreasing and eventually terminate. Does it always terminate? Let's try again, this time with $m_0 = 29$.

$$\begin{aligned}
m_0 &= 2^{2^2} + 2^{2+1} + 2^2 + 1 && = 29\\
m_1 &= 3^{3^3} + 3^{3+1} + 3^3 && \sim 8 \cdot 10^{12}
\end{aligned}$$

$$
\begin{aligned}
m_2 &= 4^{4^4} + 4^{4+1} + 4^4 - 1 = && \sim 10^{154} \\
&= 4^{4^4} + 4^{4+1} + 3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4 + 3 \\
m_3 &= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 2 && \sim 10^{2200} \\
m_4 &= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 1 && \sim 10^{36305}
\end{aligned}
$$

This does not look like it will terminate. Let's try it one more time, this time with $m_0 = 4$, to really understand what is going on.

$$
\begin{aligned}
m_0 &= 2^2 && = 4 \\
m_1 &= 3^3 - 1 = 2 \cdot 3^2 + 2 \cdot 3 + 2 && = 26 \\
m_2 &= 2 \cdot 4^2 + 2 \cdot 4 + 1 && = 41 \\
m_3 &= 2 \cdot 5^2 + 2 \cdot 5 && = 60 \\
&\vdots && \vdots \\
m_9 &= 2 \cdot 11^2 + 11 && = 253 \\
m_{10} &= 2 \cdot 12^2 + 12 - 1 = 2 \cdot 12^2 + 11 && = 299 \\
&\vdots && \vdots \\
m_{22} &= 2 \cdot 24^2 - 1 = 24^2 + 23 \cdot 24 + 23 && = 1151 \\
&\vdots && \vdots \\
m_{B-2} &= 2 \cdot B^1 && = 2B > 10^{10^8} \\
m_{B-1} &= 2 \cdot (B+1)^1 - 1 = (B+1)^1 + B && = 2B + 1 \\
m_B &= (B+2)^1 + B - 1 && = 2B + 1 \\
&\vdots && \vdots \\
m_{B+k} &= (B+k+2)^1 + B - (k+1) && = 2B + 1 \\
&\vdots && \vdots \\
m_{2B-2} &= (2B)^1 + 1 && = 2B + 1 \\
m_{2B-1} &= (2B+1)^1 && = 2B + 1 \\
m_{2B} &= (2B+2)^1 - 1 = 2B + 1 && = 2B + 1 \\
m_{2B+1} &= 2B + 1 - 1 && = 2B \\
&\vdots && \vdots \\
m_{2B+k} &= 2B + 1 - k && = 2B - (k-1) \\
&\vdots && \vdots \\
m_{3B} &= 2B + 1 - B = 1 && = B + 1 \\
&\vdots && \vdots \\
m_{4B} &= 2B + 1 - 2B = 1 && = 1 \\
m_{4B+1} &= 0 && = 0
\end{aligned}
$$

This demonstrates that when we start at $m_0 = 4$, the sequence first rises to index $B - 2$, where it attains the value $2 \cdot B^1$ (current base is B). In the next step, this is decomposed to $(B+1)^1 + B$ (this is also the maximum value). The sequence lingers here for the next B steps (each base change raises the value by 1, and this is immediately subtracted), and afterward, it begins its long descent. Finally, at index $4B+1$, it reaches zero. The exact value of B is $3 \cdot 2^{402\,653\,209} - 1$, and the length of this sequence is $4B + 2 = 3 \cdot 2^{402\,653\,211} - 2$.

Goodstein's theorem claims that a similar fate awaits every Goodstein sequence, no matter how large the starting value might be.

Theorem 3.61 (Goodstein, 1944)**.** *For every natural number m, there exists a natural number n such that $m_n = 0$.*

Definition 3.62 (Change of base function)**.** For a natural number $b \geq 2$ we define the functions

(a) *change of base function* $R_b : \omega \to \omega$ as the function that takes a natural number n and replaces each b in the hereditary base-b representation of n with $b+1$.

(b) *ω-change of base function* $R_b^\omega : \omega \to \varepsilon_0$ as the function that takes a natural number n and replaces each b in the hereditary base-b representation of n with ω.

Formally, we define $R_b(0) = R_b^\omega(0) = 0$ and for $n > 0$ expressed in base-b as

$$n = \sum_{i=0}^{k} b^i \cdot p_i$$

we let

$$R_b(n) = \sum_{i=0}^{k} (b+1)^{R_b(i)} \cdot p_i, \qquad R_b^\omega(n) = \sum_{i=0}^{k} \omega^{R_b^\omega(i)} \cdot p_i.$$

Example. For example:

- $R_2(29) = R_2(2^{2^2} + 2^{2+1} + 2^2 + 1) = 3^{3^3} + 3^{3+1} + 3^3 + 1,$
- $R_2^\omega(29) = R_2^\omega(2^{2^2} + 2^{2+1} + 2^2 + 1) = \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1$

Remark. Since each $n \in \omega$ is finite, the hereditary base-b representation of n is finite, and thus $R_b^\omega(n)$ also contains only finitely many occurrences of ω. Therefore $R_b^\omega(n) < \varepsilon_0$ for each $b \geq 2$ and $n \in \omega$.

Observation 3.63. *The terms of a Goodstein sequence starting in m could be defined as $m_{n+1} = R_{n+2}(m_n) - 1$.*

Lemma 3.64. *For every $b \geq 2$ and $n \geq 0$ it holds that $R_b^\omega(n+1) > R_b^\omega(n)$.*

Intuition. This should seem obvious, for example when $b = 2$ and $n = 13$ we get

$$\begin{aligned} R_2^\omega(13) &= R_2^\omega(2^{2+1} + 2^2 + 1) = \omega^{\omega+1} + \omega^\omega + 1 \\ R_2^\omega(13+1) &= R_2^\omega(2^{2+1} + 2^2 + 2) = \omega^{\omega+1} + \omega^\omega + \omega, \end{aligned}$$

Proof. Let $b \geq 2$ be given; we prove the claim using induction on n. If $n = 0$, then we have $R_b^\omega(1) = 1 > 0 = R_b^\omega(0)$. If $n > 0$, then let the following be the base-b expansions of n and $n+1$:

$$n = b^{c_0} \cdot p_0 + b^{c_1} \cdot p_1 + \cdots + b^{c_k} \cdot p_k$$
$$n + 1 = b^{d_0} \cdot q_0 + b^{d_1} \cdot q_1 + \cdots + b^{d_l} \cdot q_l$$

where all p_i and q_j are nonzero, and $c_0 > c_1 > \cdots > c_k$ and $d_0 > d_1 > \cdots > d_l$. Theorem 3.33, or more precisely, part two of Theorem 3.31, describes the two sufficient and necessary conditions that these expansions must satisfy in order for one of them to be greater than the other ($n + 1 > n$). Notice that the claim already holds for all exponents c_i and d_j from the induction hypothesis. Thus, when we apply R_b^ω to these expansions—change the base b to ω and apply R_b^ω to the exponents—the sufficient and necessary conditions will not be affected; therefore $R_b^\omega(n+1) > R_b^\omega(n)$. □

We can now prove Goodstein's theorem.

Proof of Theorem 3.61 (cf. [23]). Let $m = m_0$ be given. We will define a sequence of ordinals $\mu_n < \varepsilon_0$ satisfying

$$\mu_n > 0 \implies \mu_{n+1} < \mu_n \qquad \text{and} \qquad \mu_n > 0 \iff m_n > 0.$$

Since ε_0 is well-founded, it does not admit infinite strictly decreasing sequences; thus, there exists an index k at which $\mu_k = 0$, and therefore also $m_k = 0$.

Define $\mu_n \coloneqq R_{n+2}^\omega(m_n)$. Clearly $\mu_n > 0 \iff m_n > 0$, so it remains to show that μ_n is decreasing. If $\mu_n > 0$, then

$$\begin{aligned} \mu_{n+1} &= R_{n+3}^\omega(m_{n+1}) \\ &= R_{n+3}^\omega(R_{n+2}(m_n) - 1) \\ &< R_{n+3}^\omega(R_{n+2}(m_n)) \\ &= R_{n+2}^\omega(m_n) = \mu_n. \end{aligned}$$

The inequality holds by Lemma 3.64 since if $\mu_n > 0$, then $m_n > 0$ and so $R_{n+2}(m_n) \geq m_n > 0$. The equality after the inequality is a trivial property of the base change functions ($n + 2 \to n + 3 \to \omega$ is the same as $n + 2 \to \omega$). □

Extended Goodstein's theorem In the version of Goodstein sequences presented above, we started with base $b_0 = 2$ for m_0, then changed it to $b_1 = 3$ for m_1, and in general worked with base $b_n = n + 2$ for m_n. This can be generalized by considering any non-decreasing sequence $2 \leq b_0 \leq b_1 \leq \cdots$ of bases and defining the term m_{n+1} from m_n by expressing m_n in hereditary base-b_n notation, replacing each occurrence of b_n with b_{n+1}, and subtracting one. It is not hard to modify the proof above to show that this sequence still terminates.

This extended version is, in fact, the one Goodstein originally considered in [12], and he proved that it is equivalent to the claim that ε_0 is well-founded. Recall that we mentioned in Section 3.4.4 that PA cannot justify the well-foundedness of ε_0. Does that mean that Goodstein showed that the extended Goodstein's theorem cannot be proved in PA? Not quite; the theorem cannot even be stated in

PA as it is impossible to formalize. The theory lacks the expressive power to represent arbitrary infinite sequences.

The simple version we considered can however be formalized in PA easily (as we are working with only one specific sequence), and it seems much "tamer" than the extended version. For a long time it was unknown whether the simple version could be proved without using tools beyond the reach of PA (such as the well-foundedness of ε_0). Almost forty years later, Kirby and Paris [18] showed that no such finitary proof is possible.

Theorem 3.65 (Kirby–Paris, 1982)**.** *Goodstein's theorem is true, but unprovable in Peano arithmetic.*

Intuition. Even though for every fixed natural number m we have

$$\mathsf{PA} \vdash (\exists n)(\overline{m}_n = 0),$$

where $\overline{m} = S(S(\dots S(0)\dots))$ repeated m times, (for each fixed m, PA can verify the finite descent of the Goodstein sequence by explicit computation, showing that it terminates after some finite number n of steps), it also holds that

$$\mathsf{PA} \nvdash (\forall m)(\exists n)(m_n = 0).$$

Proof sketch. Kirby and Paris started with a statement of the form

$$(\forall a)(\forall b)(\exists c)\,\varphi(a,b,c)$$

that was known to be independent[10] in PA. Then they constructed a nonstandard model M of PA containing a nonstandard $b_0 \in M$ such that

$$M \models \neg(\exists y)\,\varphi(1, b_0, y).$$

That is, the above-mentioned independent statement does not hold in this model. To finish the proof, they showed that if Goodstein's theorem could be proved in PA, then there would exist a (very large) number $e \in M$ (satisfying $m_e = 0$ for some carefully chosen m defined using the nonstandard b_0) such that $\varphi(1, b_0, e)$ holds (inside M), which is a contradiction. □

This result should be surprising. As shown in [23], PA is equivalent to the theory of finite sets; that is, ZFC with the axiom of infinity replaced by the axiom "there are no limit ordinals." From this, one can prove that all sets are finite. The Kirby–Paris theorem asserts that accepting an axiom about infinite sets changes what we can prove about finite sets.

3.5.2 Kirby–Paris Hydra Game

There are different versions of the Hydra game; the one we will focus on was presented by Kirby and Paris in [18], (1982).

A *hydra* is a finite rooted tree, usually drawn with the root at the bottom. A *head* of the hydra is a leaf together with its attached edge. A *battle* between Hercules and a given hydra is divided into stages, starting at stage one. During stage n, Hercules chops off one head of the hydra. The hydra then grows n new "head segments" in the following manner:

[10] The formula $\varphi(a,b,c)$ talks about certain large ordinals $\alpha < \varepsilon_0$, utilizing the fact that these numbers have a finite hereditary Cantor normal form, and can therefore be formalized in PA.

- From the node that used to be attached to the head which was just chopped off, move along the length of one edge towards the root; that is, move to the grandparent of the chopped off node.
- From this node, sprout n replicas of that part of the hydra (after decapitation), which is "above" the edge just traveled.
- If the head just chopped off was attached to the root, no new head is grown.

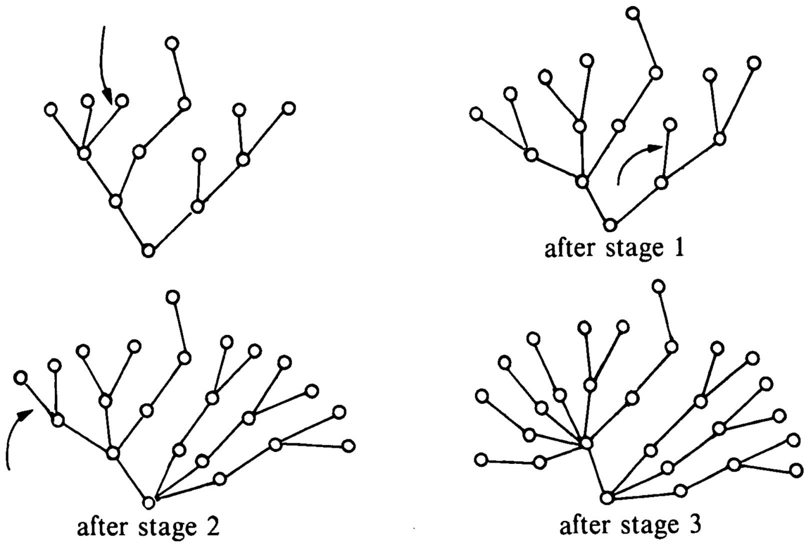

Figure 4: The Hydra game, as presented by Kirby and Paris [18], if at each stage Hercules decides to chop off the head marked with an arrow.

Hercules *wins* if after some finite number of stages, the hydra is dead; that is, nothing is left of the hydra but its root. A *strategy* is a function which determines for Hercules which head to chop off at each stage of any battle. A strategy is *winning* if it ensures that Hercules wins against every hydra. It is not hard to show that a winning strategy exists (for example, always targeting one of the highest positioned heads). More surprisingly, Hercules cannot help winning:

Theorem 3.66. *Every strategy is a winning strategy.*

Proof sketch. The idea of the proof is very similar to that of Goodstein's theorem. We create a sequence of ordinal numbers $\mu_n < \varepsilon_0$ that is strictly decreasing, and its elements are positive if and only if the hydra is still alive. Since ε_0 is well-founded, it admits no infinite decreasing chains; thus, the sequence must eventually terminate, and the hydra will die with it.

Suppose we assign to each hydra an ordinal $\alpha < \varepsilon_0$. For any strategy σ, we can define a function $H_\sigma(\alpha, n) : \varepsilon_0 \times \omega \to \varepsilon_0$ that maps the ordinal of the hydra represented by α, together with a stage number n, to the ordinal of the hydra that results from Hercules chopping off the head of α as specified by the strategy σ at stage $n - 1$ (that is, n new "heads" are grown).

To prove the theorem, one only has to show that for any strategy σ, any $0 < \alpha < \varepsilon_0$, and any $n \in \omega$, it holds that $H_\sigma(\alpha, n) < \alpha$. $\square$

Exercise 15. Finish the proof by assigning to each hydra an ordinal α such that $\alpha > 0$ if and only if the hydra is still alive, and $H_\sigma(\alpha, n) < \alpha$.

Kirby and Paris then proved a second independence statement.

Theorem 3.67 (Kirby–Paris, 1982)**.** *The statement "every recursive strategy is a winning strategy" is not provable in Peano arithmetic.*

Remark. "Recursive strategy" means any strategy that can be implemented by an algorithm. The reason why the theorem does not simply say "every strategy" is that it is not possible to represent an arbitrary infinite object in PA. The restriction to recursive strategies ensures that each one of them can be formalized by a (finite) Turing machine.

3.5.3 Fundamental Sequences

Definition 3.68. The *fundamental sequence* of a countable limit ordinal α is an increasing sequence $\alpha[0] < \alpha[1] < \dots$ such that $\alpha = \sup\{\alpha[n] \mid n < \omega\}$.

Remark. Note that Lemma 3.36 implies that no uncountable ordinal can have a fundamental sequence.

It is clear from the definition that the same ordinal α may have multiple fundamental sequences, but there is usually a "standard" one.

Definition 3.69. The following is a common assignment of fundamental sequences to limit ordinals $\alpha \le \varepsilon_0$ (the first fixed point of $\xi \mapsto \omega^\xi$). Corollary 3.32 claims that every $\alpha < \varepsilon_0$ can be expressed as

$$\alpha = \omega^{\gamma_0} + \omega^{\gamma_1} + \dots + \omega^{\gamma_k},$$

where $\gamma_0 \ge \gamma_1 \ge \dots \ge \gamma_k$ (by expanding the Cantor normal form of α). Since $\alpha < \varepsilon_0$, each exponent γ_i satisfies $\gamma_i < \alpha$. Also, note that since α is limit, we have that $\gamma_k \ne 0$. We define $\alpha[n]$ inductively as

(i) $\omega^{\gamma+1}[n] := \omega^\gamma \cdot (n+1)$,

(ii) $\omega^\gamma[n] := \omega^{\gamma[n]}$ for limit ordinals γ

(iii) $(\omega^{\gamma_0} + \dots + \omega^{\gamma_k})[n] := \omega^{\gamma_0} + \dots + (\omega^{\gamma_k}[n])$ for $\gamma_0 \ge \dots \ge \gamma_k$,

(iv) $\varepsilon_0[n] = \gamma_n$, where $\gamma_0 = 1$ and $\gamma_{n+1} = \omega^{\gamma_n}$. That is, $\gamma_n = {}^n\omega$ is a ω-tower of height n.

Example. Some simple fundamental sequences are

- $\omega[n] = \omega^0 \cdot (n+1) = n+1$,
- $\omega^\omega[n] = \omega^{\omega[n]} = \omega^{n+1}$,
- $\omega^{\omega+5}[n] = \omega^{\omega+4} \cdot (n+1)$,
- $(\omega^{\omega+2} + \omega^\omega \cdot 5)[n] = \omega^{\omega+2} + \omega^\omega \cdot 4 + (\omega^\omega[n]) = \omega^{\omega+2} + \omega^\omega \cdot 4 + \omega^{n+1}$

Fundamental sequences can also be assigned to limit ordinals larger than ε_0, but this becomes much more complicated. The next common approach after Cantor normal form is to utilize the Veblen function described in Section 3.3.2. One can show that all ordinals $\alpha < \Gamma_0$ have a unique normal form in terms of Veblen functions; for details, see Section 8 of [9]. This normal form can then be utilized to define fundamental sequences for all limit ordinals $\alpha < \Gamma_0$. If you are interested in how the formulas look, see [28].

3.5.4 Fast-Growing Hierarchy

One of the use cases of fundamental sequences is the definition of hierarchies of functions $f_\alpha : \omega \to \omega$, where each function grows faster than the previous one.

Definition 3.70 (Fast-growing hierarchy[11])**.** For ordinals $\alpha \le \varepsilon_0$, we define functions from natural numbers to natural numbers $f_\alpha : \omega \to \omega$ as follows:

(i) $f_0(n) \coloneqq n + 1$,

(ii) $f_{\alpha+1}(n) \coloneqq f_\alpha^n(n) = f_\alpha(f_\alpha(\dots f_\alpha(n)\dots))$, where f_α is composed n times,

(iii) $f_\alpha(n) \coloneqq f_{\alpha[n]}(n)$ for limit ordinals α.

Remark. Martin Löb and Stanley Wainer introduced this hierarchy in 1970s as a generalization of the Grzegorczyk hierarchy, which only considered $\alpha < \omega$.

Remark. The ordinal ε_0 is not important in the definition; we could use any other large countable ordinal μ if we had fundamental sequences for all limit ordinals $\alpha \le \mu$. Also, note that the values of the functions f_α may differ based on the chosen fundamental sequences for limit ordinals. The idea is that the values will be asymptotically the same.

Example. Some fast-growing hierarchy functions are

- $f_1(n) = 2n, \quad f_2(n) = 2^n \cdot n$,
- $f_3(n) > {}^n2$, where ${}^n2$ is a 2-tower of height n; this is called *tetration.*
- $f_\omega(n) > A(n, n)$, where A is the *Ackermann function.*[12]

Definition 3.71. A function $f : \omega \to \omega$ is

(a) *total* if it is defined for every $n \in \omega$; *partial* otherwise,

(b) *recursive* (or *computable*) if there exists an algorithm that for any given input n halts precisely when $f(n)$ is defined and outputs $f(n)$,

(c) *primitive recursive* if there exists an algorithm that does not use recursion (it only uses loops and conditions) and for any given input n halts precisely when $f(n)$ is defined and outputs $f(n)$.

[11] Numberphile has a VIDEO where they explore the growth rates of some extremely fast-growing functions (Graham's iteration and the TREE sequence).

[12] Computerphile has a VIDEO about the Ackermann function.

It is important to realize that all the functions f_α defined above are recursive, as for each of them, there is a straightforward algorithm to compute the value $f_\alpha(n)$ for any n by following α down to smaller ordinals using the fundamental sequences $\alpha[n]$. Even if the ordinal α is infinite, a Turing machine set to compute f_α would eventually find $f_\alpha(n)$ in a finite amount of time.

We have already mentioned that we can define fundamental sequences for ordinals $\alpha \leq \Gamma_0$ (and there are ways to go beyond), and the functions f_α would remain recursive as long as we have a well-defined, recursive method for choosing the sequence $\alpha[n]$. This raises the question: "How far can we go?" Eventually, it must become impossible to choose $\alpha[n]$ in a recursive manner because there are only countably many recursive functions (each corresponds to a Turing machine, which can be represented as a finite sequence of natural numbers, and there are only countably many of those), while there are uncountably many countable ordinals, as we have shown in Section 3.3.3.

Therefore, at some point, we will reach a countable "non-recursive" ordinal Λ for which f_Λ can no longer be recursive. The first such ordinal is called the *Church–Kleene ordinal* and it is denoted by ω_1^{CK}. All ordinals $\alpha < \omega_1^{\mathrm{CK}}$ are *recursive*; that means there exists a recursive well-ordering $<_\alpha$ of ω with type α. How big is the Church–Kleene ordinal? It is certainly much, much larger than Γ_0, but it is still countable, so nothing compared to ω_1.

Ackermann Function

The previously mentioned Ackermann function is a famous total recursive function that is not primitive recursive. It is defined as follows.

Algorithm 1: Ackermann function $A(m,n)$

```
Function Ackermann(m,n):
    if m = 0 then
        return n + 1;
    else
        if n = 0 then
            return Ackermann(m − 1,1);
        else
            return Ackermann(m − 1,Ackermann(m,n − 1));
```

Exercise 16. Convince yourself that the algorithm above always halts.

Definition 3.72. We say that a function $f : \omega \to \omega$ dominates a function $g : \omega \to \omega$ if, for all sufficiently large n, we have $f(n) > g(n)$.

The fast-growing hierarchy is sometimes referred to as the Grzegorczyk hierarchy due to the following theorem.

Theorem 3.73 (Grzegorczyk, 1953)**.** *Any primitive recursive function is eventually dominated by some f_n for $n \in \omega$.*

This means that the functions f_n for $n \in \omega$ measure the growth rates of all primitive recursive functions.

Corollary 3.74. *Since the Ackermann function is not primitive recursive, $A(n,n)$ dominates all of these. It can also be shown that $A(n,n)$ is dominated by $f_\omega(n)$, so the Ackermann function lies at the very edge between primitive and non-primitive recursiveness.*

Hardy Hierarchy

Definition 3.75 (Hardy hierarchy, 1904)**.** For ordinals $\alpha \leq \varepsilon_0$, we define functions from natural numbers to natural numbers $H_\alpha : \omega \to \omega$ as follows:

(i) $H_0(n) \coloneqq n$,

(ii) $H_{\alpha+1}(n) \coloneqq H_\alpha(n+1)$,

(iii) $H_\alpha(n) \coloneqq H_{\alpha[n]}(n)$ for limit ordinals α.

Example. Some simple Hardy functions are

- $H_1(n) = H_0(n+1) = n+1$,
- $H_k(n) = n+k$,
- $H_\omega(n) = H_{\omega[n]}(n) = H_{n+1}(n) = n + (n+1) = 2n+1$.

Exercise 17. Try computing $H_{\omega+\omega}(n), H_{\omega\cdot k}(n)$ and $H_{\omega\cdot\omega}$.

The Hardy hierarchy seems to be much slower than the fast-growing hierarchy; they are in fact related by $f_\alpha \sim H_{\omega^\alpha}$ for all $\alpha \leq \varepsilon_0$. However, notice that the Hardy hierarchy "catches up" at $\alpha = \varepsilon_0$ (since $\varepsilon_0 = \omega^{\varepsilon_0}$) in the sense that

$$f_{\varepsilon_0}(n-1) \leq H_{\varepsilon_0}(n) \leq f_{\varepsilon_0}(n+1).$$

This means that the two hierarchies can often be treated as equal.

Theorem 3.76 (Schwichtenberg–Wainer, c. 1972)**.** *The total recursive functions that can be proved total by Peano arithmetic are exactly those that are eventually dominated by some f_α (or equivalently some H_α) for $\alpha < \varepsilon_0$.*

Remark. By "PA can prove that f is total," we mean that PA can prove that the algorithm defining f always terminates.

Corollary 3.77. PA *cannot prove that neither f_{ε_0} nor H_{ε_0} are total.*

This is a "computational" counterpart of Gentzen's consistency proof we saw in Section 3.4.4. Gentzen showed that PA cannot prove that ε_0 is well-founded. Intuitively, it should not be able to prove that the recursive definition of H_{ε_0} is total (that every recursive call eventually reaches the base case H_0).

Connection to Goodstein's Theorem

Definition 3.78 (Goodstein function)**.** The function $\mathcal{G} : \omega \to \omega$ mapping each natural number m to the length of the Goodstein sequence starting in m is called the *Goodstein function.* That is, if n is the first index where $m_n = 0$, then $\mathcal{G}(m) = n + 1$ (since we index from zero).

Example. The first few values of the Goodstein function are

- $\mathcal{G}(1) = 2, \quad \mathcal{G}(2) = 4, \quad \mathcal{G}(3) = 6,$
- $\mathcal{G}(4) = 3 \cdot 2^{402\,653\,211} - 2 > 10^{10^8} = 10^{100\,000\,000},$
- $\mathcal{G}(5) > 10^{10^{10^{10\,000}}}, \quad \mathcal{G}(12) >$ Graham's number.[13]

Observation 3.79. *The Goodstein function $\mathcal{G}$ dominates f_α for every $\alpha < \varepsilon_0$.*

Proof. Notice that we could reformulate Goodstein's theorem as follows: the function $\mathcal{G} : \omega \to \omega$ is total. Now, if $\mathcal{G}$ was dominated by some f_α, then PA could prove that $\mathcal{G}$ is total, thus proving Goodstein's theorem. □

Corollary 3.80. *The Goodstein function is not primitive recursive since it dominates the Ackerman function $A(n,n) \sim f_\omega(n)$.*

Theorem 3.81 (Cichon, 1983)**.** *The Goodstein function is dominated by H_{ε_0} (and thus also by f_{ε_0}). More precisely,*

$$\mathcal{G}(n) = H_{R_2^\omega(n+1)}(1) - 1,$$

where $R_2^\omega(n)$ is the result of writing n in hereditary base-2 notation and then replacing all 2s with ω (as we did when proving Goodstein's theorem).

It might not be immediately obvious that this is dominated by H_{ε_0}, but consider what happens when for example $n = 15 = 2^{2^2} - 1$. Then

$$\mathcal{G}(15) = H_{R_2^\omega(15+1)}(1) - 1 = H_{\omega^{\omega^\omega}}(1) - 1, \qquad H_{\varepsilon_0}(15) = H_{{}^{15}\omega}(15),$$

where ${}^{15}\omega$ denotes the ω-tower of height 15.

Theorem 3.82 (Caicedo, 2007)**.** *If $n = 2^{m_1} + 2^{m_2} + \cdots + 2^{m_k}$ is the base-2 representation of n with $m_1 > m_2 > \cdots > m_k$, then*

$$\mathcal{G}(n) = f_{\alpha_1}\big(f_{\alpha_2}(\ldots f_{\alpha_k}(3)\ldots)\big) - 2,$$

where $\alpha_i = R_2^\omega(m_i)$.

4 Cardinal Numbers

Similar to how ordinal numbers represent order types of well-ordered sets, *cardinal numbers* represent *sizes* of well-ordered sets. Historically, these were two different classes of abstraction, but for us, cardinal numbers will be a special type of ordinal numbers.

[13] Numberphile has a VIDEO where Ron Graham himself explains the Graham's number.

4.1 Cardinals as Sizes of Well-Ordered Sets

We know that sizes of sets can be compared using injective mappings and that the relation $x \approx y$ is an equivalence on the class of all sets. We would now like to take the step from only *comparing* the sizes of sets to *quantifying* them. The question is, how do we represent classes

$$\{y \mid y \approx x\}$$

of all sets with the same size as x? We would like to define a mapping that assigns to each set x a set $|x|$ so that for any two sets x and y, we would have

$$x \approx y \iff |x| = |y|. \tag{4.1}$$

It would be ideal if, in addition, $x \approx |x|$. If such a mapping exists, then we call the set $|x|$ the *cardinality* of the set x.

It is fairly easy to define the cardinalities of certain classes of sets. For example, if x is finite, then there is a unique natural number n such that $x \approx n$, and we can set $|x| := n$. Similarly, if x is countable, then $|x| := \omega$. We are beginning to see a pattern: if x can be well-ordered, choose $|x|$ as the order type of one of its orderings; but which one? In order for (4.1) to hold, we have to choose $|x|$ as the *least* such order type. This is the key to defining cardinal numbers.

Definition 4.1 (Cardinal numbers)**.** An ordinal number κ is a *cardinal number* if for all ordinals $\alpha < \kappa$ there is no injection $\kappa \to \alpha$. Equivalently, if

$$\alpha < \kappa \implies \alpha \prec \kappa.$$

We denote the *class of all cardinal numbers* by Cn. (Cn $\subset$ On).

We say that a cardinal number κ is the *cardinality* of a set x, and we write $|x| = \kappa$, if there exists a bijection $x \to \kappa$; in other words, if $x \approx \kappa$.

As for notation, we will usually denote cardinal numbers using letters from the middle of the Greek alphabet: $\kappa, \lambda, \mu, \nu \ldots$

Observation 4.2. *For cardinal numbers κ and λ it holds that*

(a) $\kappa < \lambda \iff \kappa \prec \lambda$,

(b) $\kappa = \lambda \iff \kappa \approx \lambda$,

Observation 4.3. *If the sets x and y have their cardinalities defined, then*

(a) $x \approx y \iff |x| = |y|$,

(b) $x \approx |x|$.

Observation 4.4. *The cardinality $|x|$ is defined $\iff$ x can be well-ordered.*

Proof. Our earlier discussion and the way we defined cardinals show that if x can be well-ordered, then $|x|$ is defined (as the least order type of the well-orderings of x). On the other hand, if $|x| = \kappa$ is defined, then we can well-order x by inheriting the order of κ. $\square$

This demonstrates that if we want to have $|x|$ defined for every set x, then we need to accept the axiom of choice (which allows us to well-order any set).

Remark. There actually is a way to define a map assigning to every set x a set $|x|$ such that (a) holds *without* using the axiom of choice. This is achieved via the axiom of foundation,[14] but we lose property (b). A. Lévy (1969) showed that without either the axiom of foundation or the axiom of choice, it is impossible to define a map assigning each set x a set $|x|$ such that (a) holds. D. Pincus (1974) showed that without the axiom of choice, there exists no mapping satisfying both (a) and (b) for all sets x.

Example. Some basic properties of Cn are:

- every $n \in \omega$ and ω are cardinal numbers,
- if $\alpha \geq \omega$, then $\alpha + 1$ is not a cardinal number,
- thus every cardinal number $\kappa \geq \omega$ is a limit ordinal number,
- but not every limit ordinal number is a cardinal number — for example, we have $\omega + \omega > \omega$, but $\omega + \omega \approx \omega$, so $\omega + \omega$ is not a cardinal number,
- by the same logic, if $\alpha > \omega$ is countable, then it is not a cardinal number,
- the first uncountable ordinal ω_1 we encountered in Section 3.3.3 is a cardinal number since every $\alpha < \omega_1$ is countable.

Lemma 4.5. *If* $A \subseteq \mathrm{Cn}$ *is a set of cardinal numbers, then* $\sup(A) = \bigcup A$ *is also a cardinal number. In other words, the class* $\mathrm{Cn} \subset \mathrm{On}$ *is closed.*

Proof. Since A is a set of ordinals, according to Lemma 3.4, $\sigma \coloneqq \sup(A) = \bigcup A$ is an ordinal number. We need to show the implication

$$\alpha < \sigma \implies \alpha \prec \sigma.$$

Let $\alpha < \sigma$. Then $\alpha \in \sigma = \bigcup A$, so $\alpha \in \kappa$ for some cardinal $\kappa \in A$. Because κ is a cardinal number and $\alpha < \kappa$, it must be that $\alpha \prec \kappa$. Furthermore, because $\kappa \in A$, we have $\kappa \subseteq \bigcup A = \sigma$ and so $\kappa \preceq \sigma$. The Cantor–Bernstein theorem together with $\alpha \prec \kappa \preceq \sigma$ implies that $\alpha \prec \sigma$. □

Theorem 4.6. *For every cardinal there exists a larger cardinal.*

Proof. Suppose for contradiction that κ is the largest cardinal. Then, for every ordinal $\alpha \geq \kappa$, there exists a bijection $\alpha \to \kappa$, so κ can be well-ordered according to the type α. Corollary 2.12 implies that if r_α, r_β are well-orderings of κ with types $\alpha \neq \beta$, then $r_\alpha \neq r_\beta$ because α and β are not isomorphic.

Notice that every well-ordering of κ is a subset of $\kappa \times \kappa$, so an element of $\mathcal{P}(\kappa \times \kappa)$. For $\alpha \geq \kappa$, denote by $R_\alpha \in \mathcal{P}(\mathcal{P}(\kappa \times \kappa))$ the set of all well-orderings of κ with type α. Finally, notice that for $\alpha \neq \beta$ we have $R_\alpha \neq R_\beta$, so we can construct an injection $\alpha \mapsto R_\alpha$ that maps the proper class $\mathrm{On} \setminus \kappa$ into the set $\mathcal{P}(\mathcal{P}(\kappa \times \kappa))$, which contradicts the axiom schema of replacement. □

[14] Briefly, this is done by showing that the axiom of foundation implies that the relation $\in$ is well-founded on V, allowing us to define a hierarchy of sets $V_0 \coloneqq \varnothing$, $V_{\alpha+1} \coloneqq \mathcal{P}(V_\alpha)$ and $V_\lambda \coloneqq \bigcup_{\alpha<\lambda} V_\alpha$ for limit λ, such that $V = \bigcup\{V_\alpha \mid \alpha \in \mathrm{On}\}$. Now, for any set x denote $\varrho(x) \coloneqq \min\{\alpha \mid x \subseteq V_\alpha\}$ and define $|x|$ as the set $\{y \mid y \approx x \wedge (\forall z)\big(z \approx x \Rightarrow \varrho(y) \leq \varrho(z)\big)\}$.

Remark. Alternatively, notice that if κ is a cardinal, then the Hartogs number (see Theorem 3.45) of κ is also a cardinal, and it is larger than κ.

Corollary 4.7. *The class of all cardinal numbers* Cn *is a proper class.*

Proof. If it were a set, then by Lemma 4.5, its supremum would be the largest cardinal number, which is impossible. $\square$

Definition 4.8. The *successor* of a cardinal κ is the smallest cardinal larger than κ, and we denote it by κ^+. Furthermore, we say that κ is the *predecessor* of κ^+. Finally, $\lambda > 0$ is a *limit* cardinal if it has no predecessor.

4.2 Infinite Cardinals $\aleph_\alpha$

Finite cardinals are not very interesting since they are the same as the natural numbers. We are interested in the cardinalities of infinite sets. We have shown that the class $\mathrm{Cn} \setminus \omega$ is closed and proper. By Lemma 3.12 there exists a unique bijective normal ordinal function $\aleph : \mathrm{On} \to \mathrm{Cn} \setminus \omega$ enumerating the cardinals that measure the sizes of infinite (well-orderable) sets. Cantor introduced the symbol $\aleph$ ("aleph"), the first letter of the Hebrew alphabet, to denote this function.

Definition 4.9. The unique normal function mapping On onto the class of all infinite cardinals is denoted by $\aleph$, and its values $\aleph(\alpha)$ are denoted by $\aleph_\alpha$.

The smallest infinite cardinal number ω (the size of countable sets) is denoted by $\aleph_0$ and read as "aleph null."

Observation 4.10. *These aleph numbers are exactly the omega numbers we have discovered in Section 3.3.3. That is,* $\aleph_\alpha = \omega_\alpha$ *and*

$$\aleph_\alpha = \{\beta \in \mathrm{On} \mid |\beta| < \aleph_\alpha\}.$$

The cardinal $\aleph_\alpha$ *is the first ordinal number with cardinality* $\aleph_\alpha$*.*

Remark. Historically, ordinals and cardinals were not concrete sets but abstract concepts: ordinals described well-ordering types, while cardinals measured size. This distinction led to the development of two parallel notation systems, ω_α and $\aleph_\alpha$. Von Neumann's 1923 set-theoretic definition of ordinals unified these ideas by providing canonical representatives for ordinal types. Today, we often write ω_α when considering the cardinal $\aleph_\alpha$ viewed as an ordinal with its well-order.

Observation 4.11. $\aleph_0 = \omega$ *is a limit cardinal, and for* $\alpha > 0$ *we have that*

$$\aleph_\alpha \textit{ is a limit cardinal} \iff \alpha \textit{ is a limit ordinal.}$$

Proof. We use the fact that a cardinal is limit when it is not isolated and prove the statement by contraposition. The claim follows from the simple observation that when $\alpha = \beta + 1$ is an isolated ordinal, then $\aleph_\alpha = \aleph_{\beta+1} = \aleph_\beta^+$. And when $\aleph_\alpha = \aleph_\beta^+$ is an isolated cardinal, then $\alpha = \beta + 1$. $\square$

Observation 4.12. *If* α *is an ordinal and* ξ *is a limit ordinal, then*

(a) $\alpha \leq \aleph_\alpha$, ... $\aleph$ *is a normal function*

(b) *there exist ordinals* α *such that* $\alpha = \aleph_\alpha$, *...see Theorem 3.14*

(c) ω_α *is a limit ordinal,* *...it is an infinite cardinal*

(d) $\aleph_\xi = \sup\{\aleph_\alpha \mid \alpha < \xi\}$ *...$\aleph$ is a normal function*

Theorem 4.13. *For every ordinal* α, *it holds that* $|\aleph_\alpha \times \aleph_\alpha| = \aleph_\alpha$.

Corollary 4.14. *If* x *can be well-ordered, then* $x \times x \approx x$. *By induction, this also holds for any finite Cartesian product* $x \times \cdots \times x \approx x$.

To prove this theorem, we first define a suitable well-ordering of $\aleph_\alpha \times \aleph_\alpha$.

Definition 4.15. For ordinals α and β we define the *maximo-lexicographical* ordering of the set $\alpha \times \beta$ as

$$(\alpha_1,\beta_1) \sqsubset (\alpha_2,\beta_2) \iff \begin{cases} \max\{\alpha_1,\beta_1\} < \max\{\alpha_2,\beta_2\}, \text{ or} \\ \max\{\alpha_1,\beta_1\} = \max\{\alpha_2,\beta_2\} \wedge \alpha_1 < \alpha_2, \text{ or} \\ \max\{\alpha_1,\beta_1\} = \max\{\alpha_2,\beta_2\} \wedge \alpha_1 = \alpha_2 \wedge \beta_1 < \beta_2. \end{cases}$$

Intuition. Picture the product $\alpha \times \beta$ as a grid whose horizontal axis is indexed by α and whose vertical axis is indexed by β. Every point (α_i,β_j) sits in a "right-angle" band determined by the value $\max\{\alpha_i,\beta_j\}$. These bands consist of all points whose coordinates share the same maximum, and the bands themselves move outward from the origin. The ordering $\sqsubset$ simply compares points by the outward distance of the bands they belong to. Once two points lie in the same band, we break ties lexicographically. First, compare the α-coordinates. If those agree, compare the β-coordinates. Thus, inside each strip, the ordering "runs along the top" from left to right, and only then "climbs upward" on the right edge. It might be helpful to draw this on a piece of paper.

Exercise 18. Prove that the ordering defined above is a well-ordering.

Exercise 19. Prove that $\omega \times \omega \approx \omega$ by showing that $(\omega \times \omega, \sqsubset)$ and $(\omega, <)$ are order-isomorphic, using the result of Exercise 5. Although one can also show that $\omega \times \omega \approx \omega$ using prime-number encodings, this set-theoretic approach is conceptually cleaner, as it does not require arithmetic.

Remark. Similarly, Exercise 6 implies that the class $\mathrm{On}\times\mathrm{On}$ well-ordered maximo-lexicographically is isomorphic to On. Under this well-ordering, each $\alpha \times \alpha$ is an initial segment of $\mathrm{On} \times \mathrm{On}$; the induced well-ordering of $\alpha \times \alpha$ is called the *canonical well-ordering* of $\alpha \times \alpha$.

Proof of Theorem 4.13. By transfinite induction on α. If $\alpha = 0$, then we have the countable case $\omega \times \omega \approx \omega$, which holds by the previous exercise. Suppose that $\alpha > 0$ and consider the maximo-lexicographical ordering of $\omega_\alpha \times \omega_\alpha$. Since this is a well-ordering, it is isomorphic to a unique ordinal η, and we claim that $\eta = \omega_\alpha$. Clearly $\omega_\alpha \preceq \omega_\alpha \times \omega_\alpha \approx \eta$. This implies that $\omega_\alpha \leq \eta$ since ω_α is a cardinal number. Suppose for contradiction that $\omega_\alpha < \eta$; that is, $\omega_\alpha = (\leftarrow, \omega_\alpha) \subset \eta$ is isomorphic to an initial segment $(\leftarrow, (\gamma, \delta))$ of $(\omega_\alpha \times \omega_\alpha, \sqsubset)$, where $(\gamma, \delta) \in \omega_\alpha \times \omega_\alpha$.

Let $\xi = \max\{\gamma,\delta\}+1$ and notice that $(\leftarrow, (\gamma,\delta)) \subseteq \xi \times \xi$. Since ω_α is a cardinal and $\xi < \omega_\alpha$, there exists $\beta < \alpha$ such that $|\xi| = \omega_\beta$. By the induction hypothesis,

$$|\xi \times \xi| = |\omega_\beta \times \omega_\beta| = \omega_\beta < \omega_\alpha.$$

This is a contradiction since ω_α is isomorphic to an initial segment of $\xi \times \xi$, and we would have $\omega_\alpha < \omega_\alpha$. $\square$

Theorem 4.16. $\mathsf{AC} \iff$ *For every infinite set x, we have $x \times x \approx x$.*

Proof sketch. The direction '$\Rightarrow$' is easy; just well-order x and use Theorem 4.13.

The reverse implication is harder, and we only sketch the proof. We show that if A is an infinite set satisfying $A \times A \approx A$, then we can well-order A, implying the well-ordering principle. Let $j : A \times A \to A$ be a bijection, and for each $a \in A$ define $C_a \coloneqq \{j(a,t) \,|\, t \in A\}$. Notice that the family $\{C_a \,|\, a \in A\}$ partitions A into "A many" copies of A. In other words, the sets C_a are pairwise disjoint and satisfy $C_a \approx A$ and $\bigcup_{a\in A} C_a = A$.

Let η be the Hartogs number (see Theorem 3.45) of A; that is, the first ordinal that does not inject into A. The idea of the proof is to try to use transfinite recursion to build a family of well-ordered subsets $A_\alpha \subseteq A$ for each $\alpha < \eta$ such that the order type of A_α is α, and the sets A_α are pairwise disjoint. Informally, we do this by placing each A_α into the slot reserved by some C_a. Since $C_a \approx A$, the set A_α can always "fit" into C_a, and the sets C_a are pairwise disjoint. Suppose the transfinite recursion succeeds, and we define all pairwise disjoint A_α with isomorphisms $j_\alpha : \alpha \to A_\alpha$. Then we can define an injection $\eta \setminus \{\varnothing\} \to A$ as $\alpha \mapsto j_\alpha(0)$, contradicting the choice of η.

Hence, the transfinite recursion must "fail," and at some step $\alpha < \eta$, we will have already used up all slots C_a. That is, for every $\beta < \alpha$, there is a slot C_{a_β} containing A_β and $\bigcup_{\beta<\alpha} C_{a_\beta} = A$. Then we define a bijection $\alpha \to A$ as $\beta \mapsto a_\beta$, allowing us to well-order A. $\square$

Theorem 4.17. *For any infinite cardinal κ it holds that*

(a) $\kappa \prec \{x \,|\, x \subseteq \kappa\} = \mathcal{P}(\kappa)$,

(b) $\kappa \approx \{x \,|\, x \subseteq \kappa \text{ is finite}\}$.

Proof. The first claim follows from Cantor's theorem. We will prove the second claim in Theorem 4.70 while assuming AC. A proof without the use of AC can be found in §4 Chapter II of [1]. $\square$

Definition 4.18. Let κ and λ be cardinals. We define cardinal numbers

(a) $\kappa + \lambda \coloneqq |(\{0\} \times \kappa) \cup (\{1\} \times \lambda)|$,

(b) $\kappa \cdot \lambda = |\lambda \times \kappa|$.

In other words, $\kappa + \lambda$ and $\kappa \cdot \lambda$ are cardinal numbers, which represent the size of the set on the right side of the equation, in contrast to ordinal addition and multiplication, which express the order type of the same set when ordered lexicographically. If we want to highlight this difference, we talk about *cardinal* addition and multiplication.

Observation 4.19. *If A and B are disjoint sets with $|A|$ and $|B|$ defined, then*

$$|A \cup B| = |A| + |B|, \qquad |A \times B| = |A| \cdot |B|.$$

Observation 4.20. *Cardinal addition and multiplication are associative, commutative, and distributive. Additionally, when restricted to ω, they are the same as the corresponding ordinal operations.*

Recall that ordinal addition and multiplication are associative; however, in general, they are not commutative or right-distributive. This is because they have to "keep track" of the underlying orderings.

Lemma 4.21. *If κ and λ are cardinals, and at least one of them is infinite, then $\kappa + \lambda = \max\{\kappa,\lambda\}$. If, in addition, they are nonzero, then $\kappa \cdot \lambda = \max\{\kappa,\lambda\}$.*

Proof. Denote $\mu \coloneqq \max\{\kappa,\lambda\}$. To show addition, consider

$$\begin{aligned} \mu \preceq \kappa + \lambda &\approx (\{0\} \times \kappa) \cup (\{1\} \times \lambda) \\ &\preceq (\{0\} \times \mu) \cup (\{1\} \times \mu) = 2 \times \mu \preceq \mu \times \mu \approx \mu. \end{aligned}$$

where the last '$\approx$' follows from Theorem 4.13. The Cantor–Bernstein theorem now implies that $\kappa + \lambda \approx \mu$, so $\kappa + \lambda = \mu$ (because they are cardinals).

Similarly for multiplication: $\mu \preceq \kappa \cdot \lambda \approx \lambda \times \kappa \preceq \mu \times \mu \approx \mu$. □

Corollary 4.22. *For any ordinals α and β, it holds that*

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \max\{\aleph_\alpha,\aleph_\beta\}.$$

Corollary 4.23. *If A is an infinite set and $|B| < |A|$, then $|A \setminus B| = |A|$.*

Proof. WLOG assume $B \subseteq A$. Since the sets $A \setminus B$ and B are disjoint, we have

$$|A| = |A \setminus B| + |B|.$$

Since $|A| > |B|$ and $|A| = \max\{|A \setminus B|, |B|\}$, we must have $|A \setminus B| = |A|$. □

Lemma 4.24 (AC)**.** *For any set S, we have $|\bigcup S| \leq |S| \cdot \sup\{|A| \mid A \in S\}$.*

Proof. Let $\kappa = |S|$ and $\lambda = \sup\{|A| \mid A \in S\}$. Since $\kappa \approx S$, we can enumerate the elements of S as $S = \{A_\alpha \mid \alpha < \kappa\}$. Moreover, each A_α injects into λ; hence, we can *choose* an injection $j_\alpha : A_\alpha \to \lambda$. For an element $a \in \bigcup S$, define

$$\alpha_a \coloneqq \min\{\alpha < \kappa \mid a \in A_\alpha\}.$$

This number indicates in which A_α does a first appear in. Notice that more elements $a \in \bigcup S$ can have the same number α_a, but that $j_{\alpha_a}(a)$ uniquely identifies a among these elements (since j_{α_a} is injective). This allows us to define an injection $g : \bigcup S \to \kappa \times \lambda$ as $a \mapsto (\alpha_a, j_{\alpha_a}(a))$. □

Corollary 4.25 (AC)**.** *The union of a collection of $\aleph_\alpha$ sets, each of cardinality at most $\aleph_\alpha$, has cardinality at most $\aleph_\alpha$. If, in addition, they are non-empty and disjoint, then the union has cardinality exactly $\aleph_\alpha$.*

4.3 Cofinality and Inaccessible Cardinals

Recall the pigeonhole principle for $\omega = \aleph_0$, which says that ω cannot be partitioned into a finite number of finite sets. Thus, if $A = \bigcup\{A_i \mid i \in I\}$ is countably infinite, then either I or one of A_i has to be countably infinite. How does this generalize to higher cardinals?

Definition 4.26. Let $(X, \leq_R)$ be a partially ordered set. We say that $Y \subseteq X$ is *cofinal* in X (or is a *cofinal subset* of X) with respect to $\leq_R$ if every $x \in X$ is bounded by some $y \in Y$; (that is, $x \leq y$).

Observation 4.27. *If Y is cofinal in X, then it contains all maximal elements of X. Moreover, the relation "to be cofinal in" is transitive.*

Example. If X has a maximum x, then $\{x\}$ is the smallest cofinal subset of X.

We are usually interested in cofinality in the context of limit ordinals.

Observation 4.28. *If α is an ordinal, $A \subseteq \alpha$, and*

(i) *$\alpha = \beta + 1$ is isolated, then A is cofinal in $\alpha \iff \beta \in A$,*

(ii) *α is limit, then A is cofinal in $\alpha \iff \sup(A) = \alpha$.*

Definition 4.29 (Cofinality)**.** The *cofinality* of a limit ordinal α, denoted by $\mathrm{cf}(\alpha)$, is the least ordinal β that is the order type of some $A \subseteq \alpha$ cofinal in α:

$$\mathrm{cf}(\alpha) \coloneqq \min\{\mathrm{otp}(A) \mid A \subseteq \alpha \wedge \sup(A) = \alpha\}.$$

Observation 4.30. $\mathrm{cf}(\alpha)$ *is the least ordinal β such that there is an increasing β-sequence $\langle \alpha_\xi \mid \xi < \beta \rangle$ with limit α. Hence* $\mathrm{cf}(\alpha)$ *is always a limit ordinal.*

Lemma 4.31. *For every (limit) ordinal α, we have*

$$\mathrm{cf}(\mathrm{cf}(\alpha)) = \mathrm{cf}(\alpha).$$

Proof. Let $\beta = \mathrm{cf}(\alpha)$ and $\gamma = \mathrm{cf}(\mathrm{cf}(\alpha))$, so $\gamma \leq \beta$. From the definition of $\mathrm{cf}(\alpha)$, there exists a cofinal subset $A \subseteq \alpha$ and an increasing function (isomorphism) $f : \beta \to A$, and a cofinal subset $B \subseteq \beta$ and an increasing function (isomorphism) $g : \gamma \to B$. Notice that the map $g \circ f : \gamma \to A$ is an isomorphism of γ and a cofinal subset of α, thus (by the definition of $\mathrm{cf}(\alpha)$), $\gamma \geq \mathrm{cf}(\alpha) = \beta$. $\square$

Lemma 4.32. *The cofinality of any (limit) ordinal α is always an infinite cardinal. More precisely, it is the length of the shortest sequence with limit α:*

$$\mathrm{cf}(\alpha) = \min\{|A| \mid A \subseteq \alpha \wedge \sup(A) = \alpha\}.$$

Proof. We know that $\mathrm{cf}(\alpha)$ is an infinite ordinal. If $\mathrm{cf}(\alpha)$ were not a cardinal number, then there would be a cardinal $\kappa < \mathrm{cf}(\alpha)$ such that $\kappa = |\,\mathrm{cf}(\alpha)|$, and a bijection $f : \kappa \to \mathrm{cf}(\alpha)$. Since $\mathrm{cf}(\alpha)$ is the cofinality of α, there exists a cofinal subset $A \subseteq \alpha$ with order type $\mathrm{cf}(\alpha)$, and an isomorphism $h : \mathrm{cf}(\alpha) \to A$. The idea is to use f to skip some terms of the $\mathrm{cf}(\alpha)$-sequence defined by h.

We define a non-decreasing function $g : \kappa \to A$ as

$$g : \beta \mapsto h\big(\sup\{f(\delta) \mid \delta < \beta\}\big).$$

Notice that this is well-defined: suppose there is some $\beta \in \kappa$ such that

$$\sup\{f(\delta) \mid \delta < \beta\} = \sup(f[\beta]) = \mathrm{cf}(\alpha),$$

then $f[\beta]$ is a cofinal subset of $\mathrm{cf}(\alpha)$, and according to the previous lemma, $f[\beta]$ has order type $\mathrm{cf}(\alpha)$. Hence, there is a bijection between $\beta < \kappa$ and $\mathrm{cf}(\alpha)$, contradicting the fact that $\kappa = |\,\mathrm{cf}(\alpha)|$ is a cardinal.

Clearly $\mathrm{Rng}(g)$ is a cofinal subset of A (and therefore also of α). We claim that it has order type $\gamma \leq \kappa < \mathrm{cf}(\alpha)$, which is a contradiction. Indeed, the function $i : \mathrm{Rng}(g) \to \kappa$ defined as

$$a \mapsto \min\{\beta \in \kappa \mid g(\beta) = a\}$$

is injective and increasing (since g is non-decreasing), and thus an isomorphism of $\mathrm{Rng}(g)$ and $x := \mathrm{Rng}(i) \subseteq \kappa$. Therefore $\mathrm{Rng}(g)$ and x have the same order type γ, and Lemma 3.3 states that $\gamma \leq \kappa$ because $x \subseteq \kappa$. □

Corollary 4.33. *For any (limit) ordinal α it holds that $\omega \leq \mathrm{cf}(\alpha) \leq |\alpha|$.*

Example. Some cofinalities we already know:

- $\mathrm{cf}(\omega) = \mathrm{cf}(\omega + \omega) = \mathrm{cf}(\omega \cdot \omega) = \mathrm{cf}(\omega^\omega) = \mathrm{cf}(\varepsilon_0) = \mathrm{cf}(\Gamma_0) = \omega$,
- in general, $\mathrm{cf}(\alpha) = \omega$ for countable (limit) α, because $\omega \leq \mathrm{cf}(\alpha) \leq |\alpha| = \omega$,
- $\mathrm{cf}(\aleph_\omega) = \omega$, since $\aleph_\omega = \sup\{\aleph_\alpha \mid \alpha < \omega\}$,
- $\mathsf{AC}_\omega \implies \mathrm{cf}(\omega_1) = \omega_1$, as Lemma 3.36 implies $\mathrm{cf}(\omega_1) \geq \omega_1$.

Definition 4.34. An infinite cardinal number $\kappa = \aleph_\alpha$ is a

(a) *regular cardinal* if $\mathrm{cf}(\kappa) = \kappa$,

(b) *singular cardinal* if $\mathrm{cf}(\kappa) < \kappa$.

Intuition. If ω_α is a regular cardinal, then it is *almost* closed on taking suprema. As long as the length of the sequence is less than $\aleph_\alpha$, the limit will never reach ω_α. However, if ω_α is singular, then it is possible to reach it from bellow via a shorter sequence.

We have already seen that $\aleph_0 = \omega$ is regular and $\aleph_\omega$ singular. Furthermore, Lemma 4.31 implies that the cofinality of any (limit) ordinal α is always a regular cardinal number. The question is: are there any regular cardinals besides $\aleph_0$? If we assume the axiom of countable choice, then ω_1 is also regular. But what if we are working in bare ZF?

M. Gitik (1979) showed that the statement "$\aleph_0$ is the only regular cardinal" cannot be disproved in ZF. In other words, there exists a model of ZF where *every* limit ordinal α has a cofinal subset of size $\aleph_0$, and there exists a sequence $(\alpha_n)_{n<\omega}$ with supremum α. In particular, this implies that Lemma 3.36 cannot be proved in ZF.

Theorem 4.35. *An infinite cardinal number κ is singular $\iff$ there exists a set X such that $\kappa = \bigcup X$, where $|X| < \kappa$ and $|x| < \kappa$ for all $x \in X$.*

Proof. '$\Rightarrow$' Let $X \subseteq \kappa$ be a cofinal subset of κ with order type $\mathrm{cf}(\kappa) < \kappa$. Then $\kappa = \bigcup X$ because $\kappa = \sup(X)$, and $\sup(X) = \bigcup X$ since X is a set of ordinals. Additionally, all $\alpha \in X$ satisfy $|\alpha| < \kappa$ because κ is a cardinal.

'$\Leftarrow$' Assume that $\kappa = \bigcup X$ where $|X| < \kappa$ and all $x \in X$ have $|x| < \kappa$. Note that the cardinalities of all these sets are well-defined since they inherit a well-order from κ. If there exists $x \in X$ such that x is cofinal in κ, then $\mathrm{cf}(\kappa) \le |x| < \kappa$. If none of $x \in X$ are cofinal in κ (and thus $\sup(x) < \kappa$) define $S := \{\sup(x) \mid x \in X\}$, and notice that S is cofinal in $\bigcup X = \kappa$. Furthermore, S is a set of ordinals, so it can be well-ordered and has cardinality $|S| \le |X|$. From this, we have that $\mathrm{cf}(\kappa) \le |S| \le |X| < \kappa$. □

Corollary 4.36 (Pigeonhole principle for cardinals)**.** *If κ is a regular cardinal and $\kappa = \bigcup X$, where $|X| < \kappa$, then there exists $x \in X$ such that $|x| = \kappa$.*

If we further assume the axiom of choice, every set will have a defined cardinality, and we can make a more general claim.

Corollary 4.37 (Pigeonhole principle for infinite sets; AC)**.** *If S is an infinite set with regular cardinality $|S|$, and $S = \bigcup X$, where $|X| < |S|$, then there exists $x \in X$ such that $|x| = |S|$.*

Theorem 4.38 (AC)**.** *Every infinite isolated cardinal $\aleph_{\alpha+1}$ is regular. Thus all infinite singular cardinals are limit.*

Proof. If it were singular, then by Theorem 4.35 it would be the supremum of at most $\aleph_\alpha$ sets of cardinality at most $\aleph_\alpha$. But Corollary 4.25 implies that such a supremum must have cardinality at most $\aleph_\alpha$ (since suprema and unions of ordinals are the same). □

Exercise 20 (AC)**.** Every infinite singular cardinal κ is the supremum of $\mathrm{cf}(\kappa)$ regular cardinals.

Theorem 4.39. *If $\aleph_\alpha > \aleph_0$ is a limit cardinal (α a limit ordinal), then*

$$\mathrm{cf}(\aleph_\alpha) = \mathrm{cf}(\alpha).$$

Proof. The cardinal $\aleph_\alpha$ is defined as the limit of the sequence $\{\aleph_\beta \mid \beta < \alpha\}$. The claim follows from the observation that the set $\{\aleph_\beta \mid \beta \in A\}$ is cofinal in $\aleph_\alpha$ if and only if A is a cofinal subset of α. This allows us to skip some terms of the sequence; in fact, we only need $\mathrm{cf}(\alpha)$ many terms. □

We know that in ZFC, every infinite isolated cardinal is regular, but we have not seen any regular limit cardinals besides $\aleph_0$. Let's try different limit ordinal indices α and check whether $\aleph_\alpha$ is regular or not.

- $\mathrm{cf}(\aleph_\omega) = \mathrm{cf}(\aleph_{\varepsilon_0}) = \mathrm{cf}(\aleph_{\Gamma_0}) = \omega$,
- $\mathrm{cf}(\aleph_\alpha) = \omega$ for any countable (limit) α,
- $\mathrm{cf}(\aleph_{\omega_1}) = \mathrm{cf}(\omega_1) = \aleph_1$,
- $\mathrm{cf}(\aleph_{\omega_{\alpha+1}}) = \mathrm{cf}(\omega_{\alpha+1}) = \aleph_{\alpha+1}$ for any isolated cardinal $\omega_{\alpha+1}$.

This clearly is not working. Notice that Theorem 4.39 implies that if $\aleph_\alpha > \aleph_0$ is a regular limit cardinal, then

$$\mathrm{cf}(\aleph_\alpha) = \mathrm{cf}(\alpha) = \aleph_\alpha.$$

Since $\mathrm{cf}(\alpha) \leq \alpha$ and $\aleph_\alpha \geq \alpha$, we conclude that $\alpha = \aleph_\alpha$. Hence, every regular limit cardinal larger than $\aleph_0$ has to be a fixed point of the aleph function.

Theorem 3.14 allows us to construct fixed points of $\aleph$ quite easily; for example, the first fixed point is the limit of the sequence

$$\kappa_0 = 0,\, \kappa_{n+1} = \aleph_{\kappa_n} \qquad \longrightarrow \qquad 0, \aleph_0, \aleph_{\aleph_0}, \aleph_{\aleph_{\aleph_0}}, \aleph_{\aleph_{\aleph_{\aleph_0}}}, \ldots$$

Let us denote it κ_0 and notice that it is singular since $\mathrm{cf}(\kappa_0) = \omega$. Similarly, any fixed point constructed using this method will have cofinality ω. Theorem 3.14 allows these fixed points to be arbitrarily large, so we can summarize this as:

Lemma 4.40. *There are arbitrarily large singular cardinals $\aleph_\alpha$ such that $\aleph_\alpha = \alpha$.*

The question is: can a fixed point of $\aleph$ be a regular cardinal? People started using the term *inaccessible* to describe these cardinals as they cannot be reached from below either by taking successor cardinals or by forming suprema of smaller cardinals. Today, we call such numbers *weakly inaccessible*.

Definition 4.41. A cardinal κ is *weakly inaccessible* if it is uncountable, regular, and a limit cardinal.

Fact 4.42. ZFC *cannot prove the existence of weakly inaccessible cardinals.*

Remark. Weakly inaccessible cardinals were first introduced by F. Hausdorff in 1908. Only much later was it shown that the existence of such cardinals cannot be proved in ZFC (provided ZFC is consistent). Therefore, it is consistent with ZFC to assume that every uncountable limit cardinal number is singular.

4.4 Large Cardinals

In this section, we assume AC everywhere, unless stated otherwise.

There are many problems in set theory that lead to the question of whether there exist cardinal numbers with certain properties. If the existence of such numbers cannot be proved in ZFC, then we call them *large cardinals*. There is an entire hierarchy of large cardinals, and the weakly inaccessible cardinals are the smallest of them all. In the following sections, we will encounter other examples of large cardinals.

The typical reason why it is impossible to prove the existence of cardinals with certain properties is that they would be large enough to imply the consistency of ZFC, contradicting Gödel's second incompleteness theorem. In this section, we will attempt to provide an explanation of why this should be true.

Definition 4.43. For a set a and class B, we define the class of all mappings from a to B as

$${}^aB \coloneqq \{f \mid f : a \to B\}.$$

Definition 4.44 (Cardinal power). For cardinals κ and λ, we define the cardinal number $\kappa^\lambda := |{}^\lambda\kappa|$, that is, the size of the set of all functions from λ to κ.

So if A and B are sets, then $|{}^A B| = |B|^{|A|}$.

Observation 4.45. *In particular, 2^λ is the cardinality of the set of all characteristic functions of subsets of λ, so $2^\lambda = |\mathcal{P}(\lambda)|$.*

This means that Cantor's theorem can be formulated in ZFC as

$$(\forall\kappa \in \mathrm{Cn}) : 2^\kappa > \kappa, \quad \text{or equivalently as} \quad (\forall\kappa \in \mathrm{Cn}) : 2^\kappa \geq \kappa^+.$$

Definition 4.46. A cardinal κ is *strongly limit* if all $\lambda < \kappa$ satisfy $2^\lambda < \kappa$.

Definition 4.47. A cardinal is *strongly inaccessible* if it is uncountable, regular, and a strongly limit cardinal.

Clearly, each strongly limit cardinal is limit, and each weakly inaccessible cardinal is strongly inaccessible. Intuitively, weakly inaccessible cardinals cannot be reached from below by taking suprema, and strongly limit cardinals cannot be reached even by using the power set operation.

4.4.1 Continuum Hypothesis

The *continuum hypothesis* CH is the statement of ZFC

$$2^{\aleph_0} = \aleph_1,$$

and was first proposed by Cantor in 1878. Later, in 1908, Hausdorff proposed the *generalized continuum hypothesis* GCH, which states

$$(\forall\alpha \in \mathrm{On}) : \ 2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

We have already encountered the continuum hypothesis at the end of Section 3.4.3, where we mentioned that Gödel (1940) showed that GCH cannot be disproved from ZFC, and Cohen (1963) showed that GCH cannot be proved in ZFC. Thus GCH is independent in ZFC. Moreover, Kruse and Rubin (1960) showed that GHC implies AC over ZF.

Observation 4.48. GHC *implies that all weakly inaccessible cardinals are strongly inaccessible.*

Proof. Let κ be weakly inaccessible and $\lambda < \kappa$ be an infinite cardinal. Then $2^\lambda = \lambda^+$ and $\lambda^+ < \kappa$ because κ is limit. □

4.4.2 Cumulative Hierarchy of Sets V_α

The *von Neumann universe* is a proper class built in stages, which are referred to as the *cumulative hierarchy of sets*. We start with the empty set, and at each stage, we collect every possible subset. We define the hierarchy recursively as follows:

$$V_0 := \varnothing, \quad V_{\alpha+1} := \mathcal{P}(V_\alpha), \quad V_\lambda := \bigcup_{\alpha<\lambda} V_\alpha, \quad WF := \bigcup_{\alpha\in\mathrm{On}} V_\alpha,$$

where λ is a limit ordinal. The class WF is the von Neumann universe.

Lemma 4.49. *For all ordinals α it holds that*

(a) V_α *is a transitive set,* ... *that is,* $x \in y \in V_\alpha \implies x \in V_\alpha$

(b) $V_\beta \subseteq V_\alpha$ *for all* $\beta < \alpha$.

Proof. By transfinite induction. Note that V_α is transitive if and only if $A \in V_\alpha$ implies $A \subseteq V_\alpha$ for all A. Both claims are clearly true for $\alpha = 0$ and for limit α (assuming the claim holds for all $\beta < \alpha$). Finally, if V_α is transitive and $A \in V_\alpha$, so $A \subseteq V_\alpha$ and $A \in V_{\alpha+1}$. Hence $V_\alpha \subseteq V_{\alpha+1}$. If $A \in V_\alpha$, then $A \subseteq V_\alpha \subseteq V_{\alpha+1}$, so $A \subseteq V_{\alpha+1}$. Thus $V_{\alpha+1}$ is transitive. □

Corollary 4.50. *WF is a transitive class since it is a union of transitive sets.*

Fact 4.51. *WF is the $\subseteq$-largest transitive class on which the relation $\in$ is well-founded. This is why WF is sometimes called the* well-founded core.

Since On is a transitive class, this implies that $\mathrm{On} \subseteq WF$, and it can be shown that for all ordinals α it holds that

$$\alpha = V_\alpha \cap \mathrm{On}.$$

Let us recall that the *axiom of foundation* is the following statement:

$$(\forall A \neq \varnothing)(\exists x \in A)(x \cap A = 0).$$

Hence, for all $y \in A$, we have $y \notin x$; in other words, x is a $\in$-minimal element of the set A. The axiom of foundation is thus the statement that the relation $\in$ is well-founded on the universal class V (each set has a minimal element). Since V is transitive, we must have $WF = V$. It is then easy to see that:

Theorem 4.52. *The following statements are equivalent:*

(1) *the axiom of foundation,*

(2) $V = WF$,

(3) *for every set x, there is some ordinal α such that $x \in V_\alpha$.*

4.4.3 Beth Numbers $\beth_\alpha$

As for the sizes of the sets V_α, clearly all V_n for $n \in \omega$ are finite, and thus $|V_\omega| = \omega$. Also notice that $|V_{\alpha+1}| = 2^{|V_\alpha|}$. The sizes of the sets V_α for infinite ordinals α are known as *beth numbers*, conventionally written as $\beth_0, \beth_1, \beth_2, \ldots,$ where $\beth$ ("beth") is the second letter of the Hebrew alphabet.

Definition 4.53. The *beth numbers* $\beth_\alpha$ are defined recursively as

(i) $\beth_0 \coloneqq \aleph_0$,

(ii) $\beth_{\alpha+1} \coloneqq 2^{\beth_\alpha}$,

(iii) $\beth_\lambda \coloneqq \sup\{\beth_\alpha \mid \alpha < \lambda\}$ for limit ordinals λ.

Observation 4.54. *It is easy to verify that:*

(a) $|V_{\omega+\alpha}| = \beth_\alpha$ *for all ordinals* α*,*

(b) $\beth_\alpha \geq \aleph_\alpha$ *for all ordinals* α*,*

(c) $\beth_\omega$ *is the first strongly limit cardinal,*

(d) CH *is equivalent to* $\beth_1 = \aleph_1$*,*

(e) GCH *is equivalent to* $(\forall\alpha \in \mathrm{On}) : \beth_\alpha = \aleph_\alpha$*.*

Observation 4.55. $\kappa \in \mathrm{Cn}$ *is strongly limit* $\iff$ $\kappa = \beth_\lambda$ *for a limit ordinal* λ*.*

Proof. Clearly $\beth_\lambda$ is a strongly limit cardinal whenever λ is a limit ordinal. Let κ be a strongly limit cardinal. Since the beth numbers are unbounded, there exists a least beth number $\beth_\alpha \geq \kappa$. If $\alpha = \beta + 1$ were isolated, then $\beth_\beta < \kappa$, but $2^{\beth_\beta} = \beth_\alpha \geq \kappa$. □

Cantor's theorem implies that $\beth_\alpha$ is a normal function. Therefore by Theorem 3.14 (iv), there is a unique bijective normal function $f : \mathrm{On} \to \{\alpha \,|\, \beth_\alpha = \alpha\}$ enumerating the fixed points of the beth function.

Observation 4.56. *If* α *is a fixed point of the beth function then* $\beth_\alpha = \aleph_\alpha$*.*

Proof. Theorem 3.15 (ii) implies that the simultaneous fixed points of $\beth$ and $\aleph$ are exactly the fixed points of $\beth$, thus if $\beth_\alpha = \alpha$, then also $\aleph_\alpha = \alpha$. □

4.4.4 Constructible Universe L_α

At the end of Section 3.4.3 we mentioned that Gödel introduced the *constructible universe* L. Its construction is similar to that of the cumulative hierarchy, but we restrict set formation to only definable subsets. Informally, denote by $\mathrm{Def}(X)$ the set of subsets of X definable by first-order formulas over $(X, \in)$ with parameters from X. The hierarchy of constructible sets is defined recursively as follows:

$$L_0 \coloneqq \varnothing, \quad L_{\alpha+1} \coloneqq \mathrm{Def}(L_\alpha), \quad L_\lambda \coloneqq \bigcup_{\alpha<\lambda} L_\alpha, \quad L \coloneqq \bigcup_{\alpha\in\mathrm{On}} L_\alpha,$$

where λ is a limit ordinal. The class L is called the *constructible universe.*

To understand $\mathrm{Def}(X)$, recall that the full power set $\mathcal{P}(X)$ contains *every* conceivable collection of elements from X. In contrast, $\mathrm{Def}(X)$ contains only those subsets that we can explicitly describe using the sets we already have. Formally, a subset $Y \subseteq X$ belongs to $\mathrm{Def}(X)$ if there is a first-order formula ϕ (possibly with parameters from X) such that $Y = \{u \in X \,|\, (X, \in) \models \phi(u)\}$. When we say the formula is evaluated "over $(X, \in)$," we mean that all quantifiers $\forall$ and $\exists$ in ϕ are restricted to range strictly over the elements of X, rather than the entire universe V (so $\forall a$ is interpreted as $\forall a \in X$).

Comparing V_α and L_α It is clear from the definition that $L_\alpha \subseteq V_\alpha$ for all α. Furthermore, for every finite n, we have

$$L_n = V_n, \qquad |L_n| = |V_n| < \omega,$$

from this also

$$L_\omega = V_\omega, \qquad |L_\omega| = |V_\omega| = \omega.$$

Then

$$|V_{\omega+1}| = |\mathcal{P}(V_\omega)| = 2^\omega,$$

but it can be shown that

$$|L_{\omega+1}| = \omega.$$

Fact 4.57. *For all ordinals α it holds that*

(a) *L_α is a transitive set,*

(b) *$L_\beta \subseteq L_\alpha$ for all $\beta < \alpha$,*

(c) *$\alpha = L_\alpha \cap \mathrm{On}$,*

(d) *$|L_\alpha| = |\alpha|$ for all $\alpha \geq \omega$.*

Corollary 4.58. *L is a transitive class and* $\mathrm{On} \subseteq L$*.*

Given any model $(M, \in^M)$ of ZF, the class L^M within this model is what we call an *inner model* — the sub-universe L^M, together with the restriction of $\in^M$ to L^M is also a model of ZF. Gödel showed that this inner model additionally satisfies AC and GCH. Hence, if ZF is consistent (has a model M), then ZFC and $\mathsf{ZFC} + \mathsf{GCH}$ are also consistent (they have a model, namely L^M). This does not contradict Gödel's second incompleteness theorem because ZF cannot prove that it has a model. This is called a *relative consistency proof.*

4.4.5 Significance of the Axiom of Infinity

We have already seen a shadow of inaccessible cardinals — the cardinal $\aleph_0$. It is not uncountable, but it is a regular and strongly limit cardinal. This is the key to understanding large cardinals. Let us denote by Inf the axiom of infinity, by $\mathsf{ZFC}^- := \mathsf{ZFC} - \mathsf{Inf}$ the theory we obtain from ZFC by removing the axiom of infinity, and by $\mathsf{ZFC}_{\mathsf{fin}} := \mathsf{ZFC}^- + \neg\mathsf{Inf}$ the theory of finite sets; that is, ZFC with the axiom of infinity replaced by its negation (there are no limit ordinals). One can prove that all sets in $\mathsf{ZFC}_{\mathsf{fin}}$ are finite, and as shown in [23], $\mathsf{ZFC}_{\mathsf{fin}}$ is equivalent to PA. For a theory T, we denote by $\mathrm{Con}(T)$ the statement that T is consistent. For a more formal definition see Section 3.4.3.

Lemma 4.59. *V_ω is a model of both* ZFC^- *and* $\mathsf{ZFC}_{\mathsf{fin}}$*.*

Proof. We need to show that the axioms of $\mathsf{ZFC}_{\mathsf{fin}}$ hold in V_ω. V_ω is the set of all hereditary finite sets — sets whose elements are finite sets, whose elements are finite sets, and so on, all the way down to the empty set. Hence $\neg\mathsf{Inf}$ holds, and the axioms of ZFC^- hold because "finite sets behave nicely":

Since the elements of V_ω are sets in the outer theory, they have to satisfy extensionality. The relation $\in$ is well-founded on V_ω (foundation), the union of finite sets is finite (pairing, union), all subsets of a finite set are finite (schema of specification), the power set of a finite set is finite (power set), and the image of a finite set under a function is finite (replacement). Since finite sets can always be well-ordered, the axiom of choice is also true. $\square$

Theorem 4.60. $\mathsf{ZFC} \vdash \mathrm{Con}(\mathsf{ZFC}^-)$*, and thus, by Gödel's second incompleteness theorem, we have the two following results:*

(a) ZFC^- *cannot prove the axiom of infinity.*

(b) *the consistency of* ZFC^- *does not imply the consistency of* ZFC*.*

Proof. Because ZFC can define V_ω and prove that $(V_\omega, \in)$ is a model of ZFC^-, it follows that ZFC proves the consistency of ZFC^-. If ZFC^- proved Inf, it would thus be able to prove its own consistency. And if $\mathrm{Con}(\mathsf{ZFC}^-)$ implied $\mathrm{Con}(\mathsf{ZFC})$, ZFC would be able to prove its own consistency, as it proves $\mathrm{Con}(\mathsf{ZFC}^-)$. $\square$

Therefore, in accepting the axiom of infinity, we have to take a leap of faith. This is in direct contrast to accepting the axiom of choice — we cannot break the consistency of ZF by adding AC, but we might be breaking the consistency of ZF^- by adding Inf; no one can know for sure. In a certain sense, the axiom of infinity significantly increases the power of the theory, allowing it to prove the consistency of its weaker version.

A little bit of history Even though ZF is a formalist, axiomatic theory, its axioms should be obviously true statements. But what proof do we have (in the real world) of the existence of infinite sets? Philosophers differentiate between two views or concepts of infinity. No matter how many natural numbers one writes down, there will always be only finitely many of them. However, it will always be possible to add a new one; hence, natural numbers are *potentially infinite.* Similarly, one can always extend a line segment, so line segments also represent a potential infinity. If someone somehow created *all* the natural numbers (it would not be possible to write down any more; all of them would be there), or created an *entire* line (impossible to extend any further), then one would create what is known as an *actual infinity.*

Philosophers have pondered for centuries the question of whether an actual infinity can exist, but the consensus has generally been negative. The best proof of actual infinity we have to this day comes from the Bohemian philosopher and theologian Bernard Bolzano, and his 1851 book "The Paradoxes of the Infinite." The idea of the proof is as follows:

Consider the set of all "truths." Let S_1 be the proposition: "There are truths." This is true: if there were no truths, reasoning would be impossible. Let S_2 be the proposition: "The proposition S_1 is true." Let S_3 be the proposition: "The proposition S_2 is true." And so on. By induction, all of the propositions S_n are true, and they represent distinct truths since they talk about different objects. Therefore, there are infinitely many truths. God, in His perfection, must see all truths; hence, He possesses the actual infinity of all truths.

Mathematicians have historically worked almost exclusively with potential infinity. The birth of set theory came in 1874 when Cantor published an article in which he proved that the real numbers are uncountable, but the algebraic numbers (roots of polynomials with integer coefficients) are countable. He discussed the infinities represented by this very carefully, not yet addressing them as actual infinity. We know from a letter he sent to Dedekind that Cantor later found Bolzano's book and used Bolzano's arguments to defend the actual infinity represented by ω and $\aleph_0$, which he explicitly introduced in his later papers.

Not everyone in the mathematical world agreed with Cantor. Namely, the German mathematician Leopold Kronecker strongly disagreed with Cantor's work and prevented Cantor from publishing in the oldest and most prestigious mathematics journal at the time. However, Kronecker was far from the only one standing against Cantor. Many famous French mathematicians, including Borel, Lebesgue, and Poincaré, were skeptical of some aspects of Cantor's set theory for a long time, before finally accepting it. Cantor's work was finally widely accepted at the beginning of the 20th century, when it was backed by Hilbert who famously said "No one shall expel us from the paradise that Cantor has created."

My main source for this history section was a lecture given by prof. Petr Vopěnka, available here: `https://www.youtube.com/watch?v=_b_rPG3bu0Y`.

4.4.6 Large Cardinal Axioms

Similar to how assuming the existence of a limit ordinal allows us to prove the consistency of ZFC^-, assuming the existence of an inaccessible cardinal allows us to prove the consistency of ZFC. To illustrate how large cardinal proofs are done, we will prove (following [17]) that the existence of strongly inaccessible cardinals cannot be proved in ZFC.

Lemma 4.61. *The following are equivalent for any uncountable cardinal κ:*

(1) *κ is strongly inaccessible,*

(2) *for every $x \in V_\kappa$, if $f : x \to \kappa$, then $\sup(\operatorname{Rng} f) < \kappa$,*

(3) *for every $\alpha < \kappa$, if $f : 2^\alpha \to \kappa$, then $\sup(\operatorname{Rng} f) < \kappa$.*

Proof. (2) $\Rightarrow$ (3): Trivial.

(3) $\Rightarrow$ (1): If κ is not regular, then there is some $\alpha < \kappa$ and $f : \alpha \to \kappa$ with $\sup(\operatorname{Rng} f) = \kappa$, this f can be trivially extended to 2^α. If κ is not strongly limit, let $\alpha < \kappa$ be the first ordinal such that $\kappa \leq 2^\alpha$. Since $\kappa \preceq 2^\alpha$, there is an injection from k into 2^α, which can be reversed; therefore, there is $f : 2^\alpha \to \kappa$ which is surjective, so $\sup(\operatorname{Rng} f) = \kappa$.

(1) $\Rightarrow$ (2): It is enough to check that (2) holds for $x = V_\alpha$ for all $\alpha < \kappa$, since if $x \in V_\kappa$, there is some $\alpha < \kappa$ such that $x \subseteq V_\alpha$ (κ is a limit ordinal). Thus it is enough to show that if $\alpha < \kappa$, then $|V_\alpha| < \kappa$, since κ is a regular cardinal. We prove this by transfinite induction on α. It clearly holds for $\alpha = 0$. Suppose it holds for α, then $|V_\alpha| = \lambda < \kappa$, so $|V_{\alpha+1}| = |\mathcal{P}(V_\alpha)| = 2^\lambda < \kappa$ since κ is strongly limit. Suppose α is limit and for all $\beta < \alpha$ we have $|V_\beta| = \lambda_\beta < \kappa$. Since $\alpha < \kappa$ and κ is regular, we have

$$|V_\alpha| = \Big|\bigcup_{\beta<\alpha} V_\beta\Big| = \sup_{\beta<\alpha} \lambda_\beta < \kappa.$$

The second equality holds thanks to Corollary 4.74. $\square$

Theorem 4.62. *If κ is strongly inaccessible, then V_κ is a model of* ZFC.

Proof. Since the elements of V_κ are sets in the outer theory, they have to satisfy Extensionality. The relation $\in$ is well-founded on V_κ (Foundation); κ is uncountable, so $\omega \in V_\kappa$ since $\kappa \subseteq V_\kappa$ (Infinity). Since κ is a limit ordinal, Power set, Union, and Pairing hold. Since Union holds and each set of the outer theory can be well-ordered (we are working in ZFC), we have for every set $x \in V_\kappa$ that $\bigcup x \in V_\kappa$ can be well-ordered, so Choice holds.

Note that Specification follows from Replacement, so it is enough to show that. Let $A \in V_\kappa$ and let φ be such that $V_\kappa \models \varphi(x,y)$ corresponds to a function with domain $\supseteq A$. We want to show that $B \coloneqq \{y \mid (\exists x \in A) : \varphi(x,y)\} \in V_\kappa$. Suppose $\varphi(x,y)$ holds and define $f(x)$ as the first ordinal α such that $y \in V_{\alpha+1}$. Then f is a function from A to κ, so by (2) in the previous lemma, f must be bounded by some $\lambda < \kappa$, so $\forall y \in B$ we have $y \in V_{\lambda+1}$, so $B \in V_{\lambda+2} \subseteq V_\kappa$. □

Remark. Using a similar argument, one can show that if κ is weakly inaccessible, then L_κ is a model of ZFC.

Denote by Inf^* the statement "there exists a strongly inaccessible cardinal." Then we can state a similar theorem about Inf^* and ZFC, as we did in the previous section about Inf and ZFC^- (see Theorem 4.60).

Corollary 4.63. ZFC *cannot prove* Inf^**, and the consistency of* ZFC *does not imply the consistency of* $\mathsf{ZFC} + \mathsf{Inf}^*$.

Proof. Direct consequence of Gödel's second incompleteness theorem, and the previous theorem's claim that $\mathsf{ZFC} + \mathsf{Inf}^* \vdash \mathrm{Con}(\mathsf{ZFC})$. □

One could now imagine some property P such that if a cardinal κ satisfies P, then V_κ is a model of $\mathsf{ZFC} + \mathsf{Inf}^*$, thus $\mathsf{ZFC} + \mathsf{Inf}^*$ cannot prove the existence of such a cardinal, and so on.

In this light, we can view large cardinals as generalizations of the axiom of infinity — by accepting the existence of a large cardinal, we significantly increase the strength of our theory; however, we might be adding some inconsistencies. Many problems in set theory reduce to "if it is consistent with ZFC to assume the existence of certain large cardinals, then it is consistent to assume that something holds," or "if a certain large cardinal exists, then something holds."

The following hierarchy displays some large cardinals you might encounter ordered by consistency strength:

$$\mathsf{Inaccessible} < \mathsf{Mahlo} < \cdots < \mathsf{Measurable} < \mathsf{Woodin} < \mathsf{Supercompact} < \cdots$$

The "gap" represented by the dots above corresponds to a massive jump in logical complexity. The cardinals on the left (*inaccessible* and *Mahlo*) are essentially just "very tall" milestones. They are consistent with the constructible universe L discussed earlier. Intuitively, L is the "minimalist" version of set theory, containing no "random" or "chaotic" sets. If we assume $V = L$ (that every set is constructible), inaccessible and Mahlo cardinals can still exist. They fit inside the tidy, rigid structure of L.

However, once we cross the gap to *measurable* cardinals, this is no longer true. These cardinals are so powerful that their existence contradicts $V = L$.

They require the universe to contain objects (specifically, complex ultrafilters or embeddings) that are too "rich" to be constructible and thus cannot be inside L. Their existence implies that the universe must be much "wider" than L, containing extra information that L cannot see.

In the missing region (the dots), we find cardinals that bridge this gap. *Weakly compact* cardinals sit at the very top of the hierarchy compatible with L. Above them, we find *Erdős* and *Ramsey* cardinals, which are defined by partition properties (a generalization of Ramsey's theorem). These are one of the first cardinals that truly break out of the constructible universe, paving the way for the measurable cardinals above them. We will encounter these "partition" cardinals in Section 5.3.

The large cardinal hierarchy is highly complex, and I suggest reading [17] as an introduction. For a more thorough exposition, see [16]. For an overview of large cardinals and their properties, see THE UPPER ATTIC of Cantor's attic.

4.5 Cardinal Arithmetic

Cardinal arithmetic is the study of powers, infinite sums, and infinite products of cardinal numbers. In this section we will always assume AC, so every set is equivalent to a unique cardinal number.

Let us first recall that if κ and λ are non-zero cardinals and at least one of them is infinite, then

$$\kappa + \lambda = \kappa \cdot \lambda = \max\{\kappa,\lambda\}.$$

Furthermore, recall that $\kappa^\lambda = |{}^\lambda\kappa|$ is the size of the set of all functions $f : \lambda \to \kappa$.

Exercise 21. Show from the definition of cardinal power that

(a) if $0 < \kappa \leq \mu$ and $\lambda \leq \nu$ then $\kappa^\lambda \leq \mu^\nu$,

(b) $\kappa^{\mu+\nu} = \kappa^\mu \cdot \kappa^\nu$,

(c) $(\kappa^\mu)^\nu = \kappa^{\mu\cdot\nu}$.

Hence cardinal power is monotone and follows the usual properties of exponents.

One can show by induction that the cardinal power of natural numbers n and m is the same as the corresponding ordinal power. It is also not difficult to determine the value κ^λ when one of the arguments is a natural number:

Theorem 4.64. *For any cardinals κ, λ and $n \in \omega$ it holds that*

(a) $0^0 = 1, \qquad \lambda > 0 \implies 0^\lambda = 0,$

(b) $\kappa^0 = 1, \qquad 1^\lambda = 1,$

(c) *if $\kappa \geq \omega$ and $n > 0$, then $\kappa^n = \kappa$,* ... *by induction from $\kappa \cdot \kappa = \kappa$*

(d) *if $\lambda \geq \omega$ and $2 \leq \kappa \leq \lambda$, then $\kappa^\lambda = 2^\lambda$.* ... *in particular $\lambda^\lambda = 2^\lambda$*

Proof. (d) holds because ${}^\lambda\kappa \subseteq \mathcal{P}(\lambda \times \kappa)$ and from our assumption $\lambda \cdot \kappa = \lambda$, so

$$2^\lambda \leq \kappa^\lambda \leq |\mathcal{P}(\lambda \times \kappa)| = 2^{\lambda\cdot\kappa} = 2^\lambda. \qquad \square$$

4.5.1 Cardinality of the Continuum

It is well-known that $2^\omega = |\mathcal{P}(\omega)| = |\mathbb{R}|$. Since $\mathbb{R}$ is sometimes referred to as "the continuum," people have started calling the cardinality of these sets the *cardinality of the continuum*, and it is commonly denoted by $\mathfrak{c} := 2^\omega = \beth_1$. Other sets with this cardinality are for example the irrational numbers (as $\mathbb{Q} \approx \omega$) and transcendental numbers (as Cantor famously proved that there are only countably many algebraic numbers).

Exercise 22. The set of all open subsets of $\mathbb{R}$ has cardinality $\mathfrak{c}$. Note that the set of all subsets of $\mathbb{R}$ has cardinality $2^\mathfrak{c}$.

Hint. Each open set is a countable disjoint unions of open intervals.

Exercise 23. The set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ has cardinality $\mathfrak{c}$. Note that the set of all real functions has (by definition) size $\mathfrak{c}^\mathfrak{c} = 2^\mathfrak{c}$.

Hint. Realize that if a continuous function f is known on the rational numbers $\mathbb{Q}$, then it is known everywhere. There is only one way to continuously fill in the gaps between the rational points.

Corollary 4.65. *There exists a real function g that intersects the graph of every continuous function $f : \mathbb{R} \to \mathbb{R}$.*

Proof. The previous exercise implies that there is a way to bijectively assign to each real number r a continuous function $f_r : \mathbb{R} \to \mathbb{R}$. We define $g(r) := f_r(r)$. □

Exercise 24. The set of all well-orderings of $\mathbb{R}$ has cardinality $2^\mathfrak{c}$. Note that the set of all binary relations on $\mathbb{R}$ has size $2^{|\mathbb{R}\times\mathbb{R}|} = 2^{\mathfrak{c}\cdot\mathfrak{c}} = 2^\mathfrak{c}$.

Hint. Consider all ordinal types of well-orders of $\mathbb{R}$, and for each order-type consider the possible ways to order $\mathbb{R}$ using this type.

4.5.2 Infinite Sums and Products

Definition 4.66 (AC)**.** The (infinite) *sum* and *product* of a collection of cardinal numbers $\langle \kappa_i \mid i \in I \rangle$ are the cardinals

$$\sum_{i\in I} \kappa_i := \Big|\bigcup_{i\in I}(\{i\} \times \kappa_i)\Big|, \qquad \prod_{i\in I} \kappa_i := \Big|\bigtimes_{i\in I} \kappa_i\Big|.$$

Exercise 25. The lexicographic ordering of $\bigtimes_{i\in\omega} \kappa_i$ is not a well-ordering.

It is easy to see that the sum and product of cardinals will not change when we permute the indices in I. If λ_i are cardinals such that $\kappa_i \le \lambda_i$ for all $i \in I$, and if $J \subseteq I$, then

$$\sum_{i\in I} \kappa_i \le \sum_{i\in I} \lambda_i, \qquad \sum_{i\in J} \kappa_i \le \sum_{i\in I} \kappa_i.$$

Furthermore, if $\langle A_i \mid i \in I \rangle$ is a collection of sets, then

$$\Big|\bigcup_{i\in I} A_i\Big| \le \sum_{i\in I} |A_i|, \qquad \Big|\bigtimes_{i\in I} A_i\Big| = \prod_{i\in I} |A_i|.$$

If, in addition, the sets are disjoint, then $\left|\bigcup A_i\right| = \sum |A_i|$.

Observation 4.67. *If $\kappa_i = \kappa$ for all $i \in I$, then $\sum \kappa_i = |I| \cdot \kappa$ and $\prod \kappa_i = \kappa^{|I|}$.*

Definition 4.68. If X is a set and λ is a cardinal, define

(a) $[X]^\lambda := \{x \subseteq X \mid |x| = \lambda\}$, ... subsets of size λ

(b) $[X]^{<\lambda} = \{x \subseteq X \mid |x| < \lambda\}$. ... subsets of size $< \lambda$.

If $n \geq k$ are natural numbers, then $[n]^k = \binom{n}{k}$, and $[X]^{<\omega}$ is the set of all finite subsets of X. Furthermore, if $\lambda > |X|$, then $[X]^{<\lambda} = \mathcal{P}(X)$ and $[X]^\lambda = \varnothing$.

Lemma 4.69 (AC)**.** *If X is an infinite set of size $|X| = \kappa$ and $\lambda \in \mathrm{Cn}$, then*

(a) $\left|[X]^\lambda\right| = \kappa^\lambda$, *if* $\lambda \leq \kappa$, ... *if $\lambda > \kappa$, then $\left|[X]^\lambda\right| = 0$*

(b) $\left|[X]^{<\lambda}\right| = \sum_{\mu<\lambda} \kappa^\mu$, *if* $\lambda \leq \kappa^+$, ... *if $\lambda \geq \kappa^+$ then $\left|[X]^{<\lambda}\right| = 2^\kappa$*

Proof. (a), WLOG assume $X = \kappa$, so we want to show $[\kappa]^\lambda \approx \kappa^\lambda$. For any set $x \in [\kappa]^\lambda$ there is a map $f \in {}^\lambda\kappa$ such that $x = \mathrm{Rng}(f)$. Using AC, we *choose* for each $x \in [\kappa]^\lambda$ such a mapping f_x. We have constructed an injection $[\kappa]^\lambda \to {}^\lambda\kappa$, showing $[\kappa]^\lambda \preceq \kappa^\lambda$. To show the other inequality, realize that every mapping $f : \lambda \to \kappa$ is an element of $[\lambda \times \kappa]^\lambda$, so $\kappa^\lambda \preceq [\lambda \times \kappa]^\lambda$. But since $\lambda \times \kappa \approx \kappa$, we have $[\lambda \times \kappa]^\lambda \approx [\kappa]^\lambda$.

(b) follows from (a) since the sets $[X]^\mu$ are disjoint for different values of μ. □

Corollary 4.70. *If X is an infinite set, then $\left|[X]^{<\omega}\right| = |X|$.*

Proof. This is a direct consequence of (b) and basic cardinal properties:

$$\left|[X]^{<\omega}\right| = \sum_{n<\omega} |X|^n = \sum_{n<\omega} |X| = \omega \cdot |X| = \max\{\omega, |X|\} = |X|. \qquad \square$$

Definition 4.71 (Weak power)**.** For cardinals κ and λ, define the cardinals

$$\kappa^{<\lambda} := \sum_{\mu<\lambda} \kappa^\mu, \qquad \kappa^{\leq\lambda} := \sum_{\mu<\lambda^+} \kappa^\mu.$$

Observation 4.72. *If κ is an infinite cardinal then $\kappa^{\leq\kappa} = \kappa^\kappa = 2^\kappa$.*

Proof. The previous lemma implies that $\kappa^{\leq\kappa} = 2^\kappa$ and $\kappa^\kappa = 2^\kappa$. □

Lemma 4.73 (Infinite sum)**.** *Let $\langle \kappa_i \mid i \in I \rangle$ be a collection of nonzero cardinals. If the set I or at least one of the cardinals κ_i is infinite, then*

$$\sum_{i\in I} \kappa_i = \max\{|I|, \sup_{i\in I} \kappa_i\}.$$

Proof. Let $\kappa = \sup \kappa_i$. Then $\sum \kappa_i \approx \bigcup(\{i\} \times \kappa_i) \subseteq |I \times \kappa| = \max\{|I|, \kappa\}$. We have shown "$\leq$". For $\sum_{i\in I} \kappa_i \geq \max\{|I|, \kappa\}$, we show that

- $\sum_{i\in I} \kappa_i \geq \sum_{i\in I} 1 = |I|$, and
- $\sum_{i\in I} \kappa_i \geq \kappa$ since $\sum_{i\in I} \kappa_i \geq \kappa_j$ for all j, so $\sum_{i\in I} \kappa_i$ is an upper bound, and κ is the least upper bound. □

Remark. Note that "$\geq$" also follows directly from Lemma 4.24.

Corollary 4.74. *If $\langle X_i \mid i \in I \rangle$ is a collection of sets such that $|X_i| = \kappa_i$, and $\sup\{\kappa_i \mid i \in I\} \geq |I|$ is an infinite cardinal, then*

$$\Big|\bigcup_{i\in I} X_i\Big| = \sup_{i\in I} \kappa_i = \sum_{i\in I} \kappa_i.$$

Proof. WLOG all X_i are nonempty. Let $\kappa = \sup \kappa_i$. The previous lemma claims that $\kappa = \sum \kappa_i$. We know that $\big|\bigcup X_i\big| \leq \sum \kappa_i$, and we also know that $\kappa_j \leq \big|\bigcup X_i\big|$ for all j, so $\big|\bigcup X_i\big|$ is an upper bound. Since κ is the least upper bound, we have $\kappa \leq \big|\bigcup X_i\big|$. $\square$

Observation 4.75. *If κ is an infinite cardinal with cofinality $\mathrm{cf}(\kappa)$, then there are cardinals $\langle \kappa_i \mid i \in \mathrm{cf}(\kappa) \rangle$ such that $\kappa_i < \kappa$ for all i and $\kappa = \sum \kappa_i$*

Proof. Let $X = \langle \alpha_i \mid i \in \mathrm{cf}(\kappa) \rangle$ be a cofinal subset of κ. Then $\bigcup \alpha_i = \kappa$, and by Corollary 4.74 $\kappa = \sup |\alpha_i|$. We let $\kappa_i = |\alpha_i|$; since $\alpha_i < \kappa$ and κ is a cardinal, we have $\kappa_i < \kappa$. $\square$

Corollary 4.76. *A cardinal κ is singular $\iff$ there are cardinals $\lambda < \kappa$ and $\langle \kappa_i \mid i \in \lambda \rangle$ such that $\kappa_i < \kappa$ for all i and $\kappa = \sum \kappa_i$.*

Proof. If such cardinals exist, then $\kappa = \sup\{\kappa_i \mid i < \lambda\}$, meaning that κ has a cofinal subset of size $\lambda < \kappa$. The previous observation implies the other direction. $\square$

Remark. This is a cardinal arithmetic counterpart of Theorem 4.35.

4.5.3 König's Inequality

Lemma 4.77. *If $\langle \kappa_i \mid i \in I \rangle$ is a collection of cardinals $\kappa_i \geq 2$, then*

$$\sum \kappa_i \leq \prod \kappa_i$$

Proof. If $|I| = 0$, then $\sum = 0$ and $\prod = 1$ ($\bigtimes \kappa_i$ is the set of some mappings from the empty set, and there is always the empty mapping). If $|I| = 1$, then $\sum = \prod$. If $|I| = 2$, consider cardinals κ_0 and κ_1. We want to find bijective mappings of κ_0 and κ_1 to disjoint sets $A_0, A_1 \subseteq \kappa_0 \times \kappa_1$. Imagine $\kappa_0 \times \kappa_1$ as a grid with horizontal axis κ_0 and vertical axis κ_1. We map κ_0 to the first row, and κ_1 to the first column — but since they overlap at (0,0), we map one of the zeroes (for example $0 \in \kappa_1$) to (1,1).

If $|I| \geq 3$, map $\kappa_i \setminus \{0\}$ to the "i-th edge of the box", that is the "vector" $(0,\ldots,0,*,0,\ldots 0)$, where $*$ is a wildcard at index i. And map the zero $0 \in \kappa_i$ to the "point" $(1,\ldots,1,0,1,\ldots,1)$, where the 0 is at index i. The formal definition of the mappings $\kappa_i \to \bigtimes \kappa_i$ is left to the reader. $\square$

Theorem 4.78 (König's inequality)**.** *If $I \neq \varnothing$ and $\kappa_i < \lambda_i$ for all $i \in I$, then*

$$\sum \kappa_i < \prod \lambda_i.$$

König's inequality is a generalization of Cantor's theorem. If $\kappa_i = 1$ and $\lambda_i = 2$ for all i, then König's inequality implies $|I| < 2^{|I|}$.

Proof. Notice that if $\lambda_j = 1$, then $\kappa_j = 0$; so if we remove j from I, the sum and product will not change. Hence, WLOG all $\lambda_i \geq 2$. By the previous lemma:

$$\sum \kappa_i \leq \sum \lambda_i \leq \prod \lambda_i.$$

We need to show strictness. Similar to how one can prove Cantor's theorem, we use the diagonal method. For contradiction, let $\sum \kappa_i = \prod \lambda_i$. Then there is a bijection between $\sum \kappa_i$ and $\prod \lambda_i$, and it induces a collection $\langle X_i \,|\, i \in I \rangle$ of disjoint sets X_i such that $|X_i| = \kappa_i$ and $\bigcup X_i = \bigtimes \lambda_i$. Note that the elements of the cartesian product of a collection of sets $\bigtimes \lambda_i$ are functions $f : I \to \bigcup \lambda_i$. Let

$$Y_i = \{f(i) \,|\, f \in X_i\} \subseteq \lambda_i.$$

We have $|Y_i| \leq |X_i| = \kappa_i < \lambda_i$, so $\lambda_i \setminus Y_i$ is nonempty. Let $g \in \bigtimes \lambda_i$ be defined as $g(i) = \min(\lambda_i \setminus Y_i)$ and notice that this g is not in any X_i, a contradiction. □

Corollary 4.79. *If $\kappa \geq 2$ and $\lambda \geq \omega$ are cardinals, then*

(a) $\mathrm{cf}(2^\lambda) > \lambda$,

(b) $\mathrm{cf}(\kappa^\lambda) > \lambda$,

(c) $\lambda^{\mathrm{cf}(\lambda)} > \lambda$.

Proof. (a) follows from (b).

(b) Let $\langle \kappa_i \,|\, i < \lambda \rangle$ be a sequence of cardinals $\kappa_i < \kappa^\lambda$. Then

$$\sup_{i \in \lambda} \kappa_i \leq \sum_{i \in \lambda} \kappa_i < \prod_{i \in \lambda} \kappa^\lambda = (\kappa^\lambda)^\lambda = \kappa^{\lambda \cdot \lambda} = \kappa^\lambda.$$

Where we used Lemma 4.73 and König's inequality for $\lambda_i = \kappa^\lambda$ for all i.

(c) Thanks to Observation 4.75 there are cardinals $\langle \kappa_i \,|\, i \in \mathrm{cf}(\lambda) \rangle$ such that $\kappa_i < \lambda$ and $\sup \kappa_i = \lambda$. We use König's inequality for $\lambda_i = \lambda$ for all i:

$$\lambda = \sup_{i \in \mathrm{cf}(\lambda)} \kappa_i \leq \sum_{i \in \mathrm{cf}(\lambda)} \kappa_i < \prod_{i \in \mathrm{cf}(\lambda)} \lambda = \lambda^{\mathrm{cf}(\lambda)}.$$ □

4.5.4 Indeterminacy of the Continuum Function $2^{\aleph_\alpha}$

Let us summarize what our findings up to this point imply about the *continuum function* $\aleph_\alpha \mapsto 2^{\aleph_\alpha}$. We have derived the following inequalities:

Proposition 4.80. *For any ordinals α and β the following holds:*

(i) $\alpha \leq \beta \implies 2^{\aleph_\alpha} \leq 2^{\aleph_\beta}$, *...from monotonicity*

(ii) $2^{\aleph_\alpha} > \aleph_\alpha$, *...Cantor's theorem*

(iii) $\mathrm{cf}(2^{\aleph_\alpha}) > \aleph_\alpha$. *...part* (a) *of Corollary 4.79*

Note that (iii) *implies* (ii) *since* $2^{\aleph_\alpha} \geq \mathrm{cf}(2^{\aleph_\alpha})$.

These results impose certain restrictions on the values of the continuum function. For instance, since $\mathrm{cf}(2^{\aleph_0}) > \aleph_0$, we know that $2^{\aleph_0}$ cannot be $\aleph_\omega$. However, beyond these basic bounds, ZFC tells us remarkably little about what these values actually *are*, particularly when $\aleph_\alpha$ is a regular cardinal.

Recall the generalized continuum hypothesis GCH is the statement

$$(\forall \alpha) : 2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

Easton (1970) showed that the above inequalities are all the axioms of ZFC imply about the values of $2^{\aleph_\alpha}$ for regular $\aleph_\alpha$. More formally:

Fact 4.81 (Easton)**.** *If $G : \mathrm{On} \to \mathrm{On}$ is a function that satisfies*

(i) $\alpha \leq \beta \implies \aleph_{G(\alpha)} \leq \aleph_{G(\beta)}$, ... *simulating* (i)

(ii) $\mathrm{cf}(\aleph_{G(\alpha)}) > \aleph_\alpha$, ... *simulating* (iii)

then it is consistent with ZFC *to assume that $2^{\aleph_\alpha} = \aleph_{G(\alpha)}$ holds for all regular cardinals $\aleph_\alpha$.*

Remark. This also means that $\mathfrak{c} = 2^{\aleph_0}$ could be arbitrarily large; it could even be an inaccessible cardinal.

Woodin (1981) showed that there is some "wiggle room" for the values of $2^{\aleph_\alpha}$ for singular $\aleph_\alpha$ as well. In particular:

Fact 4.82 (Woodin)**.** *It is consistent with* ZFC *to assume*

$$(\forall \alpha) : 2^{\aleph_\alpha} = \aleph_{\alpha+2},$$

provided that the existence of a supercompact cardinal is consistent with ZFC*.*

Finally, let us ponder whether GCH could be true for some initial segment of infinite cardinals. It can be shown from Easton's result that any reasonably defined regular cardinal $\aleph_\alpha$, for example

$$\aleph_0, \aleph_1, \aleph_{100}, \aleph_{\omega+1}, \aleph_{\omega_1+1},$$

can be the first on which GCH breaks. This means that $2^{\aleph_\alpha} > \aleph_{\alpha+1}$, but $2^{\aleph_\beta} = \aleph_{\beta+1}$ for all $\beta < \alpha$.

The question whether GCH can be violated for the first time by a singular cardinal has remained open for a long time. Silver (1974) showed that this cannot happen for singular cardinals with uncountable cofinality, and Magidor (1977) demonstrated that Silver's result probably cannot be extended to cardinals with countable cofinality. More specifically:

Fact 4.83 (Silver)**.** *If $\aleph_\alpha$ is a singular cardinal with $\mathrm{cf}(\aleph_\alpha) > \omega$, and $2^{\aleph_\beta} = \aleph_{\beta+1}$ holds for all $\beta < \alpha$, then also $2^{\aleph_\alpha} = \aleph_{\alpha+1}$.*

Fact 4.84 (Magidor)**.** *The statement*

$$(\forall n < \omega) : 2^{\aleph_n} = \aleph_{n+1} \quad \wedge \quad 2^{\aleph_\omega} = \aleph_{\omega+2}$$

is consistent with ZFC*, provided it is consistent with* ZFC *to assume the existence of a certain pair of large cardinals.*

Therefore, if the existence of these cardinals is consistent, then it is impossible to prove Silvers theorem for $\aleph_\omega$, which is a singular cardinal with cofinality ω, and Silvers theorem thus cannot be generalized for all singular cardinals.

Remark. The function $\aleph_\alpha^{\aleph_\beta}$ has also been studied extensively, and its behavior is much more complex than the behavior of $2^{\aleph_\alpha}$. We will not study its values here, but [1] contains a range of results and their simplification if we assume GCH.

5 Infinite Combinatorics

We assume AC everywhere unless stated otherwise.

FROM THIS POINT THE NOTES ARE NOT PROPERLY EDITED

5.1 König's Theorem for Infinite Trees

Definition 5.1. A *(model theoretic) tree* is a partially ordered set $(T, <_T)$ where for every $x \in T$ is the set $\{y \in T \mid y < x\}$ well-ordered by $<_T$.

Observation 5.2. *Every chain in a tree is a well-ordered subset.*

Note that the "branches" of the tree can be ordered more wildly than ω — they can by ordered by any ordinal. So there can be some "limit points" along the branches. Furthermore, a tree does not have to be connected.

Definition 5.3. A *branch* in a tree a $\subseteq$-maximal chain.

Observation 5.4. *If we accept the axiom of choice, then we can extent any chain into a branch using Zorn's lemma.*

Definition 5.5. For a given tree T and $x \in T$ we define

(a) the height of x is the ordinal type of $(\leftarrow, x)$, denoted y $H_T(x)$, or $H(x)$, or $|x|_T$. The tree is divided into levels, starting with level 0.

(b) $T_\alpha = \{x \in T \mid H_T(x) = \alpha\}$ is the αth level of T

(c) the height $H(T)$ of T is the least α for which $T_\alpha = \varnothing$. Also $H(T) = \sup\{H_T(x) + 1 \mid x \in T\}$. So it is the height in vertices, not in edges. So a tree with 2 up edges has height 3.

(d) the length of a branch is its ordinal type with respect to $<_T$. (It is always at most $H(T)$, using $H(T) = \sup\{H_T(x) + 1 \mid x \in T\}$).

(e) a cofinal branch is a branch of length $H(T)$.

Example.

- Every ordinal $(\alpha, <)$ is a tree of height α with a single branch.
- Complete X-ary tree of height α... vertices are sequences of elements of X of length $< \alpha$; that is the vertices are

$$X^{<\alpha} = \bigcup_{\beta<\alpha} {}^{\beta}X$$

and the order is that from each vertex, there are $|X|$ edges, going into vertices extending that vertex.

- Cantor's tree — complete 2-ary tree of height ω — infinite binary tree. It is a countable tree with $2^\omega = c$ branches. So we can use the branches to model the real numbers.

Question: is there a tree of height ω with no cofinal branch? Yes — consider a tree with ω branches of lengths 1,2,3,.... Then the tree has height ω, but none of the branches are infinite. But what happens if we require the tree to have a finite branching factor? (Thus there are only finitely many nodes on each level).

In general, for any infinite cardinal κ, there exists a tree if height κ with no cofinal branch and with levels of cardinality $\mathrm{cf}(\kappa)$ (by the same construction).

Theorem 5.6 (König's theorem, AC_ω)**.** *Every tree of height ω such that its every level is finite has a cofinal branch.*

Intuition. If the human species continues to live indefinitely, there will exist a person whose lineage will never die out (we consider only the male or only the female individuals).

Proof. The tree is a countable union of countable sets, which is countable (under AC_omega), so $|T| = \omega$. For $x \in T$ define $[x, \rightarrow)$ as the subtree starting at x, and $(x, \rightarrow)$ as the subtree containing only the descendants of x. Notice that

$$T = \bigcup_{x \in T_0} [x, \rightarrow)$$

Since $|T_0| < \omega$, by pigeonhole $\exists x_0 \in T_0$ such that $[x, \rightarrow)$ is infinite. In general, for $x_n \in T_n$, we have

$$[x_n, \rightarrow) = \bigcup_{y \in T_{n+1}, x <_T y} [y, \rightarrow).$$

From the axiom of choice, there eixsts a function $f : T \rightarrow T$ such that if $x \in T_n$ and $|(x, \rightarrow)| = \omega$, then $f(x) \in T_{n+1}, x < f(x)$ and $|(f(x), \rightarrow)| = \omega$. Basically, we assign each vertex an infinite branch starting at that vertex. By transfinite reductions we construct a sequence $(x_n : n < \omega)$ where $x_{n+1} = f(x_n)$, $x_n < x_{n+1}$ (and $[x, \rightarrow)$ are infinite), so $\{x_n \,|\, n < \omega\}$ is a cofinal branch. We basically start at the root, and follow an infinite branch. And then repeat this again. And again. □

Exercise 26. Derive König's theorem directly from Zorn's lemma.

Fact 5.7 (Aronszajn)**.** *There is a tree of height ω_1, where every level is countable, with no cofinal branch. (Balcar Stepanek 3.41)*

Fact 5.8. *If κ is an infinite cardinal with $\mathrm{cf}(\kappa) = \omega$ and every level is finite, then there exists a cofinal branch. (Balcar Stepanek 3.39)*

Fact 5.9. *If $\mathrm{cf}(\kappa) > \nu$ and every level has cardinality strictly smaller than ν, then there still exists a cofinal branch. But the result by Aronszajn demonstrates that if the levels are not* strictly *smaller, then it doesnt have to exist. (Balcar Stepanek 3.40). The proof requires the notion of stationary sets.*

5.2 Compactness Principle

Definition 5.10. For a collection of finite sets $\langle A_i \mid i \in I \rangle$ (for example A_is are sets of colors) we define a

1. *partial selector* of the collection as a mapping (partial coloring) f such that $\mathrm{Dom}(f) \subseteq I$ and $(\forall i \in \mathrm{Dom}(f))\, f(i) \in A_i$
2. *system of selectors* S covers finite subsets of I if for evert finite $u \subseteq I$ exists $f \in S$ such that $u \subseteq \mathrm{Dom}(f)$
3. a total sector is any mapping $g \in \bigtimes_{i \in I} A_i$. It is a filtered extension of a system of selectors S if for every finite $u \subseteq I$ exists $f \in S$ such that $f \restriction u = g \restriction u$.

Intuition. A system of selctors is a lot of short horizontal jiggly line segemnts. A filtered extensions is finding a long jiggly line, such that it is created from some short line segements. But it does not (and will not) have to be consistent with all the short ones.

Theorem 5.11 (Compactness principle, AC)**.** *If $\langle A_i \mid i \in I \rangle$ is a collection of finite sets, then every system of partial selectors covering finite subsets of I has a filtered extension.*

Example. Example for hypergraphs: $A_i = \{1,2,\dots r\}$ and S =system of good r-colorings (of vertices) of finite induced subgraphs. Then there exists a good r-coloring of H.

Corollary 5.12 (AC)**.** *Let $H = (V,E)$ be a hypergraph, where V is an arbitrary set, and $E \subseteq [V]^{<\omega}$ is a set of finite subsets of V, and $r \in \omega$. If every finite subgraph of H has a "good r-coloring," then the entire graph H has a good r-coloring.*

Remark. Here "good" means the standard notion of graph vertices coloring. But (almost?) any other property would work.

Another application of this theorem is proving the infinite Hall's theorem.

Theorem 5.13 (Infinite Hall's theorem)**.** *A collection $\langle A_i \mid i \in I \rangle$ of finite sets has an injective selector (a system of distinct representatives) $\iff$ every finite sub-collection has an injective selector (by compactness principle, EXERCISE). Now by finite Hall's theorem: $\iff$ Hall's condition is true for all finite subsets, that is $(\forall J \subseteq I)(|J| < \omega \implies \bigcup_{j \in J} A_j \geq |J|)$.*

Exercise 27. What if A_is are infinite? Will Hall's theorem still be true?

Proof of Compactness principle, according to Boris Buch. We will use Zorn's lemma. Let S be a system of partial selectors covering finite subsets of I. We say that a partial selector h is compatible with S if

$$(\forall u \subseteq I)(|u| < \omega \implies (\exists f \in S)(u \subseteq \mathrm{Dom}(f) \wedge f \restriction (u \cap \mathrm{Dom}(h)) = h \restriction u))$$

On all finite sets u, there is some partial selector f such that f is defined on the entire u, and f and h agree on the elements of u that h is defined on. We say that h is compatible with f on u.

Denote by P the set of all partial selectors compatible with S, ordered by $\subseteq$, so the ordering extends the selectors. Notice that $P \neq \varnothing$ as $\varnothing \in P$. To use Zorn's lemma, we need to show that every chain $\{h_j \mid j \in J\}$ has an upper bound. We claim that the bound is $H \coloneqq \bigcup_{j\in J} h_j$. It is a mapping, it is a partial selector, and it is compatible with S. The first two should be obvious, the compatibility with S is true: ;et $u \subseteq I$ be finite, then $u \cap \mathrm{Dom}(H)$ is contained in the domain of some h_j, that is compatible with S.

By Zorn's lemma, there exists a maximal element h of P. It remains to show that h is a total selector ($\mathrm{Dom}(h) = I$). For contradiction suppose that there exists $x \in I \setminus \mathrm{Dom}(h)$. That is, h cannot be extended by any pair (x,y) where $y \in A_x$. Let $h_y = h \cup \{(x,y)\}$ (we try to color the vertex x with color y). Because h is maximal, $h_y \notin P$ and h_y is not compatible with S. Thus for all $y \in A_y$, there exists a finite witness $u_y \subseteq I$ such that h_y is not compatible with any selector from S defined on $u_y \cup \{x\}$ (negation of the formula above). We take one such witness u_y for every $y \in A_x$. Since A_x is fintie $U \coloneqq \bigcup_{y\in A_X} u_y$ is finite, and $U' \coloneqq U \cup \{x\}$ is also finite. Because $h \in P$ and U' is finite, there exists $f \in S$ defined on U' compatible with h on $U \cup \{x\}$ (note that h is not defined on the full U', only U. But then we could select $y = f(x)$ and h_y is compatible with $f \in S$, which is a contradiction (h_y were supposed to be compatible with nothing, now we can extend h, so h is not maximal). $\square$

Fact 5.14 (Hadwiger–Nelson)**.** *This is a famous problem about the chromatic number of the plane, denoted by $\chi(\mathbb{R}^2)$, defined as $\chi(G)$ where $G = (\mathbb{R}^2),\{\{x,y\} \in (\mathbb{R}^2)^2 \mid ||x-y| = 1\}$. It is fairly easy to show that $4 \leq \chi(\mathbb{R}^2) \leq 7$ (wikipedia); the lower bound was recently improved to 5. But we do not know anything more.*

There are graphs whose chromatic number depends on whether or not we assume the axiom of choice. There is a graph which has chromatic number 2 if we assume AC, and the chromatic number may be uncountable without choice.

5.3 Infinite Ramsey Theorems

Exercise 28. If $\mathbb{Q} = A_1 \cup A_2 \cup \cdots \cup A_n$ is a partition of $\mathbb{Q}$ into finitely many parts, then at least one A_i contains an order-isomorphic copy of $\mathbb{Q}$. That is, so A_i has a countable dense subset with no maximum and no minimum.

We will state Ramsey theorem about partitions of finite, or countable subsets of some given set X. Note that every partition can be interpreted as a coloring — subsets in part A_i have color i.

Definition 5.15. If Q is a partition of $[X]^n$ (n element subsets), then a set $A \subseteq X$ is homogeneous (for Q) if there is some partition $q \in Q$ such that $[A]^n \subseteq q$.

Equivalently, we can define this notion using mappings. If f is a mapping defined on $[X]^n$, then $A \subseteq X$ is homogeneous (for f) if f is constant on $[A]^n$ (all elements of $[A]^n$ have the same color)

Definition 5.16 (Partition arrow)**.** If κ,λ,μ are cardinals, and $r \in \omega$, then the expression

$$\kappa \longrightarrow (\lambda)^r_\mu$$

means that for every set X of cardinality κ, and every mapping $f : [X]^r \to \mu$ (μ-colorings of r element subsets), there exists $A \subseteq X$ of cardinality λ, that is homogeneous for f.

Example.
- "every graph on 6 vertices contains a triangle or an anti-triangle" using this notation $6 \to (3)^2_2$.
- write the classical finite graph Ramsey theorem for c colors using this notation: $(\forall n)(\forall c)(\exists N) : N \to (n)^2_c$. For hypergraphs with hyper edges of size r, we have $N \to (n)^r_c$
- pigeonhole principle for regular cardinal κ: If $\mu < \kappa$, then $\kappa \to (\kappa)^1_\mu$. For singular: $\mu < \mathrm{cf}(\kappa)$.

Theorem 5.17 (Infinite Ramsey theorem)**.** $(\forall r)(\forall c)\, \omega \to (\omega)^r_c$

Proof sketch. We proceed by induction on r. The base case $r = 1$ is the pigeonhole principle. Suppose $r \geq 2$ and χ is a coloring of r-element subsets of ω. Denote $n_0 := 0$, and consider all the subsets containing $n_0 = 0$, and define a coloring of $(r-1)$ tuples $[\omega \setminus \{n_0\}]^{r-1}$ as $\chi'(A) := \chi(A \cup \{n_0\})$. We use Ramsey theorem for $r - 1$ and get an infinite homogeneous $A_0 \subseteq \omega$ so all r-tuples from A_0 contaitning n_0 have the same color w.r.t. χ.

Now we define $n_1 := \min A_0$, and find homogeneous $A_1 \subseteq A_0$ etc. We get a sequence $n_0, n_1, n_2, \ldots$ where the color of an r-tuple is determined by its smallest element. By pigeonhole, there eixsts a subsequence n_{i_1},n_{i_2},n_{i_3},$dots$ homogeneous w.r.t. χ. □

Notice that this proof doesnt use any form of Choice. But we used the fact that ω is well-ordered. The following theorem cannot be proven in ZF (Kleinberg)

Theorem 5.18. *For every coloring of* $[X]^r$ *where* X *is an infinite set, there exists an infinite homogeneous subset.*

Fact 5.19. *Infinite Ramsey implies finite ramsey by compactness.*

Definition 5.20. A set $S \subseteq \omega$ is large if $|S| > \min(S)$.

Example. {2,5,8} is large, but {6,7,8,9} is not large.

Theorem 5.21 (Paris–Harrington 1968 myslim)**.** $(\forall n,r,c \in \omega)(\exists N \in \omega)(\forall f : [N]^r \to c)$ *there exists a large set* $A \subseteq \omega$*, which is homogeneous of* f*, and* $A \geq n$*.*

Proof. For contradiction assume that for every $N \in \omega$ there exists a coloring $f : [N]^r \to c$ with no large monochromatic subset of cardinality $\geq \mu$. We will use the compactness principle for $\langle A_i \mid i \in [\omega]^r \rangle$, where $A_i = c = \{0,1,\ldots,c-1\}$. We obtain a coloring $g : [\omega]^r \to c$ with no large monochromatic subset of cardinality $\geq \mu$. But by the infinite Ramsey theorem, there exists an infinite monochromatic (for g) subset $A \subseteq \omega$. Take the first $\max(n, \min(A) + 1)$ elements of A. This is a large monochromatic (for some big N) subset of cardinality at least n, a contradiction. □

Notice that this is a natural extension of the Ramsey theorem, with the added condition that A must be large. Paris and Harrington showed that this is not provable in PA.

Exercise 29. Prove from the infinite Ramsey theorem (assuming AC) that

- Every infinite ordered set has an infinite chain or an infinite anti-chain.
- Every infinite linearly ordered set has an infinite subset of order-type ω (infinite increasing sequence), or $\omega*$ (infinite decreasing sequence).

Hint. Find a countable subset and use two colors — one for comparable elements and the other for incomparable.

Theorem 5.22 (Sierpinsky, AC)**.** $2^\omega \not\longrightarrow (\omega_1)^2_2$.

Intuition. Not only is ω_1 not enough, but 2^ω, which could be larger without CH, is also not enough

Lemma 5.23. *No subset of* $\mathbb{R}$ *is ordered according to* ω_1 *or* ω_1^* *(the reverse order)*

Proof. For contradiction suppose there is an increasing sequence of real numbers $\{x_\alpha \mid \alpha \in \omega_1\} \subseteq \mathbb{R}$ of order type ω_1. Then for every $\alpha < \omega_1$ there exists a rational number $q \in (x_\alpha, x_{\alpha+1})$, take q_α to be the smallest from this interval according to some well-ordering of $\mathbb{Q}$. We get an injective map $\omega_1 \to \mathbb{Q}$, a contradiction. ☐

Proof of theorem. Define a coloring f of $[\mathbb{R}]^2$ (2-element subsets): let $\prec$ be a well-ordering of $\mathbb{R}$ and define $f(\{x,y\})$ as 0 if $x < y$ and $x \prec y$, and as 1 if $x < y$ and $y \prec x$. Notice that we first order the unordered pair, and then use the well-order. Every monochromatic subset A of color 0 is well-orderd with respect to $<$ (because then $<$ agrees with $\prec$), color 1 wrt $>$. If $|A| > \omega$, then the order type of A would be at least ω_1 (or ω_1^*), which contradicts the lemma. ☐

Theorem 5.24 (Sierpinsky, generalization)**.** *For every infinite cardinal* κ *we have* $2^\kappa \not\to (\kappa^+)^2_2$

Proof sketch. Real numbers can be represented using the branches of the Cantor tree (sequences of 0s and 1s of length ω). Similarly we can take the lex. ordering of sequences from ${}^\kappa 2$ (0/1 sequences of length κ) has no subset of order type κ^+. Proof in Balcar Stepanek 4.75. ☐

Theorem 5.25 (Erdos–Rado)**.** *For every infinite cardinal* κ *we have* $(2^\kappa)^+ \longrightarrow (\kappa^+)^2_2$*, and in fact*

$$(2^\kappa)^+ \longrightarrow (\kappa^+)^2_\kappa,$$

and for any $n \in \omega$ *we have*

$$(2^{2^{\cdots 2^\kappa}})^+ \longrightarrow (\kappa^+)^{n+1}_\kappa,$$

where there is a 2*-tower of height* n*.*

Lemma 5.26. *For every infinite cardinal* κ*, we have* $\kappa \not\to (\omega)^\omega_2$*.*

Proof. We define an equivalence relation $\sim$ on $[\kappa]^\omega$ as

$$X \sim Y \iff X \triangle Y \text{ is finite}$$

So the two sets of size ω differ in only finitely many elements. Where $\triangle$ is the symmetric union (we dont take the intersection). Exercise: verify that $\sim$

is an equivalence. From AC we obtain a selector g on the partition into eq. classes – assigning each class an element of that class: $g([X]_\sim) \in [X]_\sim$. We define $f : [\kappa]^\omega \to [\kappa]^\omega$ as $f(X) = g([X]_\sim)$, assigning each ω-tuple to its classes representative. Define sets $Q_0 = \{X \in [\kappa]^\omega \mid |X \triangle f(X)| is even\}$ and $Q_1 = \{X \in [\kappa]^\omega \mid |X \triangle f(X)| is odd\}$. The set $\{Q_0, Q_1\}$ forms a partition of $[\kappa]^\omega$, and we claim that none of them has an infinite (countable) homogeneous subset. If $A \subset \kappa$ and $X \subseteq A$ are countably infinite, then $X \in [A]^\omega$. If $x \in X$ and $Y := X - \{x\}$, then $X \sim Y$, so $f(X) = f(Y)$, but $|X \triangle f(X)|$ and $|Y \triangle f(Y)|$ have different parities, so X and Y both have different colors, so A is not monochromatic (or rather $[A]^\omega$ is not). $\square$

Hence Ramsey style theorems only make sense for coloring finite tuples.

Theorem 5.27 (Erdos, Dushnik, Miller)**.** *For every infinite cardinal κ, we have $\kappa \to (\kappa, \omega)^2_2$, similarly to how in the finite Ramsey theorem we might find a clique of size at least k, or an independent set of size at least l.*

Exercise 30. If κ is an infinite cardinal, then $2^\kappa \not\to (3)^2_\kappa$. Party, pairs of people chose a language the will speak with each other κ languages, we want 3 people with the same language.

Sierpinsky says that if κ is successor, then Erdos-Dushnik-Miller cannot be improved to

$$\kappa \to (\kappa, \kappa)^2_2$$

.

Definition 5.28. κ is weakly compact cardinal if κ is uncountable and $\kappa \to (\kappa)^2_2$

Fact 5.29. *The existence of such cardinals is not provable in ZFC, they are large cardinals.*

Exercise 31. If κ is weakly compact, then it is regular and strongly limit, hence strongly inaccessible. (we know limit since Sierpinsky, and strongly limit also from Sierpinsky).

Definition 5.30. κ is a Ramsey cardinal if $\kappa \to (\kappa)^{<\omega}_2$, so we are coloring all finite subsets.

Exercise 32. We dont need to say uncountable because ω is not a Ramsey cardinal, so Ramsey is stronger than Weakly compact.

Interesting question maybe: for what κ do we have $\kappa \to (\omega)^{<\omega}_2$? Explore, maybe using some limit of Erdos-Rado would work, with some smart many color encoding?

5.4 Chromatic Number of Infinite Graphs

The following is a construction by Soifer and Shelah (2003). The vertices of G are $V = \mathbb{R}$ and x,y form an edge if $|x - y| - \sqrt{2} \in \mathbb{Q}$... so lengths of edges are $\sqrt{2}$+rational number.

Proposition 5.31. *In ZFC, $\chi(G) = 2$, so G is bipartite.*

Proof. Define $S = \{q + n\sqrt{2} \mid q \in \mathbb{Q}, n \in \mathbb{Z}\}$. So we have a copy of $\mathbb{Q}$ for every integer n, shifted by $n\sqrt{2}$. Notice that for each two neighboring levels, n and $n+1$, the graph induced by these two subsets is a complete bipartite graph. So we can color all even levels red and all odd levels blue, and get a 2-coloring of $S \subsetneq \mathbb{R}$; we want to color the missing pieces. Since S is an additive subgroup of $\mathbb{R}$, we can partition $\mathbb{R}$ into shifted copies of S. Notice that every such S is a component of G and they make up G. So we again color each copy of S using two colors, the end. But we need the axiom of choice to choose a "zero" level for each copy – to decide what will be the even/odd levels. What follows is just a formalization of this idea.

Define an equivalence relation E as $xEy \iff x - y \in S$, this is just a formalization of the idea that $\mathbb{R}$ is partitioned into disjoined copies of S. Now we use the axiom of choice: from each equivalence class $[t]_E$ (from each copy of S) pick a representative and collect them into a set y. For any $t \in \mathbb{R}$, let $f(t) \in y$ be the representative of the equivalence class $[t]_E$ (of the copy of S containing t).

Now define a 2-coloring $c : \mathbb{R} \to 2$ as follows: $c(t) = 0$ if $t - f(t) = 2n\sqrt{2} + q$ for some $n \in \mathbb{Z}$ and $q \in \mathbb{Q}$, and $c(t) = 1$ if $t - f(t) = (2n+1)\sqrt{2} + q$. Note that c is a proper coloring because the graph induced by $\mathbb{Q} \cup (\mathbb{Q} + \sqrt{2})$ is complete bipartite ?? $\square$

Axiom 5.32 (LM)**.** Every subset of $\mathbb{R}$ is Lebesgue measurable.

Short introduction to measure theory The idea is that we want to measure the "volumes" of subsets of $\mathbb{R}^n$. We would like to say that the volume of the interval $[a,b]$ is $b - a$, or that the volume of a ractangle of sides x and y is $x \cdot y$. Similaraly, the volume of the disjoint union of intervals $[a_i,b_i]$ is the sum of their volumes.

Definition 5.33 (Lebesgue measure)**.** We want to be able to measure the volumes of sets in $\mathcal{E} \subseteq \mathcal{P}(\mathbb{R})$, a system of subsets of $\mathbb{R}$, containing all intervals, and closed under countable unions and countable intersections. If we only start with intervals and take the closure of there operations, we get the Borel sets. If we also add some sparse sets with measure zero, we get the measurable sets. We want a mapping $\lambda : \mathcal{E} \to \mathbb{R}_0^+ \cup \{\infty\}$ satisfying $\lambda([a,b]) = b - a$ and if $A_1, A_2, \ldots$ are pairwise disjoint then $\lambda(\bigcup A_n) = \sum \lambda(A_n)$, for the definition of this we need AC_ω. This is like an axiomatic definition, the question is if such a mapping exists.

Fact 5.34. *If $A \in \mathcal{E}$, then $\lambda(A) = \inf\{\sum_{i=1}^{\infty}(b_i - a_i) \mid A \subseteq \bigcup_{i=1}^{\infty}(a_i,b_i)\}$. So the measure of a set is the measure of the most efficient covering of A using open intervals. So for example, for any $A = \{x\}$ has measure 0, since we can take smaller and smaller open intervals approaching x. Also, closed and open intervals have the same measure. Nice introduction to measure theory is in Mathematiky ++, for the same course, written by Matousek, Kantor and Samal.*

Theorem 5.35 (Solovay)**.** *ZF $+AC_\omega+$LM is consistent (if ZF is consistent).*

Proposition 5.36. *In ZF $+AC_\omega+$LM we have $\chi(G) \geq \aleph_1$.*

Proof. If $A_1 \cup A_2 \cup \cdots$ is a partition of $\mathbb{R}$ into countably many subsets, then there exists i such that $\lambda(A_i) > 0$. So if $\chi(G) \leq \aleph_0$, then there exists a coloring

$f : \mathbb{R} \to \omega$ and exists color i such that $\lambda(f^{-1}[\{i\}]) > 0$. To get a contradiction we claim that if $A \subseteq \mathbb{R}$ of positive Lebesgue measure, then A is not an independent set of G, so A cannot be monochromatic, a contradiction, we need more colors. Proof: The fact with infimum implies another fact: there exists an interval I such that $\lambda(A \cap I)/\lambda(I) \geq 9/10$. Idea: the measure of A is concentrated somewhere, A cannot be too sparse, otherwise it would have measure zero. This is genearalized by the density theorem. Let $q \in \mathbb{Q}$ such that $q + \sqrt{2} \in (0, \frac{\lambda(I)}{10})$. Define $B := A + (q + \sqrt{2})$. Because we shifter A only a little bit, so $\lambda(B \cap I) \geq \frac{8}{10}$, and $\lambda(A \cap B \cap I) \geq \frac{7}{10} > 0$ (imagine A in a rectangle I, and shift A a bit). In particular $A \cap B \neq \varnothing$, so let $x \in A \cap B$, and notice that $x \in A$ and $y := x - (q + \sqrt{2}) \in A$, so x and y are an edge of G, with both endpoints in A, so A is not independent. $\square$

5.5 Paradoxical Decompositions

This is happening in $\mathbb{R}^3$.

Definition 5.37. Subsets $A, B \subseteq \mathbb{R}^3$ are

1. *congruent* $A \simeq B$ if B can be obtained from A by translations and rotations.
2. *mutually decomposable* using n pieces $A \overset{n}{\simeq} B$ if there exist partitions $A = A_1 \cup \cdots \cup A_n$ and $B = B_1 \cup \cdots \cup B_n$ such that $A_i \simeq B_i$ for all i. And $A \preceq^n B$ if there is $B' \subseteq B$ such that $A \simeq^n B'$.

Observation 5.38. *A,C disjoint and B,D disjoint, then*

$$(A \simeq^n B \wedge C \simeq^m D) \implies A \cup C \simeq^{n+m} B \cup D.$$

Similarly for $\preceq^n$. As for transitivity, we have

$$A \preceq^n B \preceq^m C \implies A \preceq^{n \cdot m} C,$$

similarly for $\simeq^n$. The reason is that we overlay the two partitions, which can yield up to $n \cdot m$ pieces. Draw a picture.

Proposition 5.39 (Generalized cantor Bernstein). $A \preceq^m B \preceq^n A \implies A \simeq^{m+n} B$

Proof. Similar to standard Cantor Bernstein. The goal is to find partitions $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$ such that we can map A_1 onto B_1 using m pieces and B_2 onto A_2 using n pieces. Denote by φ the mapping from A to B and by ψ the mapping from B to A.

To get the partition we desire, we can for example take a fixed point of the mapping $H : \mathcal{P}(A) \to \mathcal{P}(A)$ as $H(u) = A - \psi[B - \varphi[u]]$.

Or for $a \in A$ we consider the sequence $\psi^{-1}(a), \varphi^{-1}(\psi^{-1}(a)), \psi^{-1}(\varphi^{-1}(\psi^{-1}(a)))$. This sequence could be finite or infinite. Denote by A_L the set

$$\{a \in A \,|\, \text{the sequence of preimages is of even length}\},$$

by A_O odd length sequences, and by A_∞ the infinite sequence starting points. So $A = A_L \cup A_O \cup A_\infty$ Similarly, we partition $B = B_L \cup B_O \cup B_\infty$. Now define $A_1 = A_L \cup A_\infty$, $A_2 = A_O$ and $B_1 = B_O \cup B_\infty$, $B_2 = B_L$. This works because when we add one more point to a odd sequence we get an even sequence and vice versa, $\square$

Exercise 33. Prove $S^1 \simeq^2 S^1 \setminus \{(1{,}0)\}$ where S^1 is the unit circle. So we can fill in a missing point in the circle for free. Similarly show that $D \simeq^{n+2} D \cup n$ line segments of length 1, where D is the unit disk (we rotate the radii). So we can generate line segments for free.

Proposition 5.40. *$S \subseteq \mathbb{R}^3$ unit sphere and $D \subset S$ a countable subset, $D' := S \setminus D$. Then $S \simeq^2 D'$. So we can fill in countably many points in the sphere.*

Intuition. In the unit circle version, consider that it is missing more tha one point - then we can make sequences form each point and rotate them all at onece. But one has to be careful if the points are spaced by some multiple of pi radians.

Proof. There exists an axis disjoint from D and a rotation α about some angle such that $\alpha[D], \alpha[\alpha[D]]$ are disjoint. So every point will have its own circle missing one point. How can we pick this axis? No two points from D should rotate on each other, so for all pairs a,b from D we need to forbid all axis contained in the hyperplane orthogonal to the segemtn ab. D is countable, so we are forbiding countably many hyperplanes, so after we remove them, the measure of what we are left with is still the same as before. Hence we can just pick a random axis and with probability 1 it will work with any angle. Or we dont care about the axis but need to be careful about the angle. Now define partitions $A = D \cup \alpha[D] \cup \alpha[\alpha[D]] \cup \cdots$ and $B = S \setminus A$. Then $A \simeq \alpha[A]$ and $B \simeq B$, so $S = A \cup B \simeq^2 \alpha[A] \cup B = (A \setminus D) \cup B = D'$. □

Theorem 5.41 (Banach–Tarski)**.** *$\bar{S}, \bar{S}_n \subset \mathbb{R}^3$ disjoint unit balls (closed), then $\bar{S} \simeq^{10} \bar{S} \cup \bar{S}_1$. But it can be done using a smaller number of parts.*

The theorem relies on the following crutial fact.

Fact 5.42. *There exist rotations α by $180°$, and β by $120°$ such that (except for $\alpha^2 = \beta^2 = e$) there is no other finite composition of α and β (for example $\alpha\beta\alpha\beta^2$) is equal to the identity. In group theory terms, α and β generate a group of rotations that is isomorphic to the free product $\mathbb{Z}_2 * \mathbb{Z}_3$ (that is $2^3 = 1$ and $3^2 = 1$, but we cannot get 1 any other way). Osofsky and Adams proved that we can take any two axis (going through the origin) with angle $\theta = 45°$ between them.*

The Banach–Tarski paradox does not work in $\mathbb{R}^2$ because no rotations with similar properties exist in $\mathbb{R}^2$. But if we allow some other affine transformations (not only rotation and trabslation) that we can create a similar "paradox" already in $\mathbb{R}^2$.

Proof of theorem. Take rotations α and β from the previous fact and the group G generated by α and β. Elements of G are the identity e and rotations of the form $(\alpha)\beta^{\epsilon_1}\alpha\beta^{\epsilon_2}\alpha\beta^{\epsilon_3}\dots$ where $\epsilon_i \in \{1{,}2\}$. Think about it, these are all the compositions and (α) denotes that it might be missing. One might need to do some linear algebra to show that every element of G is a rotation about some axis (composition of two rotations is a rotation).

Denote by D the set of all intersections of the axis of the rotations from G with the ball S and note that D is countable. For every $\delta \in G$ we have $\delta[D] = D$. Indeed, if $x \in D$ then there is $\gamma \in G$ s.t. $\gamma(x) = x$ (the rotation corresponding to the axis). We define the rotation $\delta\gamma\delta^{-1}$ and notice that $\delta(x)$ is a fixed point of this rotation (hence it lies on its axis), so $\delta(x) \in D$. Similarly we consider the rotation $\delta^{-1}\gamma\delta$ with fixed point $\delta^{-1}x$ to show the other inclusion.

Denote by $D' \coloneqq S \setminus D$, by our previous observation D' has no fixed points of nontrivial elements of G. For $x \in D'$ define $S_x \coloneqq$ the orbit of x (from group theory) as $\{\gamma(x) \mid \gamma \in G\}$ that is all points where x can be mapped. Notice that these orbits form a partition of D'. Using the axiom of choice, choose one point from every orbit and collect them into a set T. Every element of D' can be uniquely expressed as $\gamma(t)$ where $\gamma \in G$ and $t \in T$ because $x \in D'$.

Define

$$A \coloneqq \{\gamma(t) \mid t \in T, \gamma = e \vee \gamma = \alpha\beta^{\epsilon_1}\alpha\beta^{\epsilon_2}\alpha\beta^{\epsilon_3}\dots\}$$

and

$$B \coloneqq \{\gamma(t) \mid t \in T, \gamma = \beta\alpha\beta^{\epsilon_1}\alpha\beta^{\epsilon_2}\dots\}$$

and

$$C \coloneqq \{\gamma(t) \mid t \in T, \gamma = \beta^2\alpha\beta^{\epsilon_1}\alpha\beta^{\epsilon_2}\dots\}$$

Clearly $D' = A \cup B \cup D$ is a partition and $A \simeq B \simeq C \simeq A$ as $B = \beta[A]$ and $C = \beta[B]$ and $A = \beta[C]$. Also, $\alpha[B \cup C] \subset A \simeq C$. So $A \preceq^1 B$ and $B \cup C \preceq^1 C$, together $A \cup B \cup C \preceq^2 B \cup C \preceq^1 A$, so $D' = A \cup B \cup C \preceq^2 A \simeq B$.

By the proposition about the countable subset of a sphere $D \cup D' = S \simeq^2 D' \preceq^2 A$, so $S \preceq^4 A$, similarly $S_1 \preceq^4 B$ and $S \cup S_1 \preceq^8 A \cup B$. We extend to balls $\bar{S}$,$\bar{S}_1$ and consider sets $\bar{A}$,$\bar{B}$ of radii segments $0a$,$a \in A$ and $0b$,$b \in B$. Then remove origins and used the mappings for spheres $\bar{S}_0 - 0 \cup \bar{S}_1 - 0' \preceq^8 \bar{A} \cup \bar{B} \preceq \bar{S}$. There is still a lot of points in $\bar{S}$ which are not included in $\bar{A} \cup \bar{B}$, so we can map the origins to those. We can map 0 to 0 and $0'$ into some arbitrary $y \in C$. that adds one more piece and we have $\bar{S} \cup \bar{S}_1 \preceq^9 \bar{S}$. Cantor–Bernstein together with $\bar{S} \preceq^1 \bar{S}$ give $S \cup \bar{S}_1 \simeq^{10} \bar{S}$. $\square$

Exercise 34. There exists n such that carbon atom $\simeq^n$ the Sun.

Exercise 35. Similar construction does not work in $\mathbb{R}^2$. Show that if α is a rotation by 180 and β by 120, then $(\alpha\beta)^6 = e$. And find an indentity that is satisfied by any pair of rotations α,β in $\mathbb{R}^2$.

Sources

This document serves as lecture notes for the course NMAI074 taught at MFF CUNI by doc. Kynčl. The web of the course is HERE. A significant portion of the text follows parts of the second and third chapters of [1], which is in Czech. My notes from the introductory set theory course can be found HERE, also in Czech.

If you found any mistakes or errors, please contact me at smolikj@matfyz.cz.

[1] Petr Balcar Bohuslav a Štěpánek. *Teorie množin*. Vydání 2., opravené a rozšířené. Praha: Academia, 2001. ISBN: 80-200-0470-X.

[2] Stephen Budiansky. *Journey to the Edge of Reason: The Life of Kurt Gödel*. W. W. Norton, 2021. ISBN: 9781324005452.

[3] Jakub Bulín. *NAIL062 Propositional and Predicate Logic: Lecture Notes*. 2025. URL: `https://github.com/jbulin-mff-uk/nail062/raw/main/lecture/lecture-notes/lecture-notes.pdf` (visited on 11/06/2025).

[4] Jochen Burghardt. *Matchstick representation of ordinal numbers up to ω^ω*. 2023. URL: `https://commons.wikimedia.org/wiki/File:Omega-exp-omega-normal.pdf` (visited on 11/04/2025).

[5] Jochen Burghardt. *Order Type Examples*. 2019. URL: `https://commons.wikimedia.org/wiki/File:OrderTypeExamples.pdf` (visited on 11/04/2025).

[6] Timothy Y Chow. "The consistency of arithmetic". In: *The Mathematical Intelligencer* 41.1 (2019), pp. 22–30. URL: `https://arxiv.org/pdf/1807.05641`.

[7] Walter Dean and Sean Walsh. "The prehistory of the subsystems of second-order arithmetic". In: *The Review of Symbolic Logic* 10.2 (2017), pp. 357–396. URL: `https://arxiv.org/pdf/1612.06219`.

[8] Herbert B Enderton. *Elements of set theory*. Gulf Professional Publishing, 1977, pp. 195–196.

[9] Jean H Gallier. "What's so special about Kruskal's theorem and the ordinal Γ_0? A survey of some results in proof theory". In: *Annals of pure and applied logic* 53.3 (1991), pp. 199–260. URL: `https://www.cis.upenn.edu/~jean/kruskal.pdf`.

[10] Gina Garcia Tarrach. "The Axiom of Choice and its implications in mathematics". MA thesis. Universitat de Barcelona, 2017. URL: `https://hdl.handle.net/2445/121981`.

[11] Kurt Gödel. "On Formally Undecidable Propositions of Principia Mathematica and Related Systems I". In: *Kurt Gödel: Collected Works: Volume I Publications 1929-1936*. OUP Oxford, 1986, pp. 144–195. ISBN: 0-19-503964-5.

[12] Reuben Louis Goodstein. "On the restricted ordinal theorem". In: *The Journal of Symbolic Logic* 9.2 (1944), pp. 33–41. DOI: `10.2307/2268019`.

[13] Gro-Tsen and IkamusumeFan. *A graphical "matchstick" representation of the ordinal ω^2*. 2015. URL: `https://commons.wikimedia.org/wiki/File:Ordinal_ww.svg` (visited on 11/04/2025).

[14] Joel David Hamkins. *Transfinite recursion as a fundamental principle in set theory*. Accessed: 2025-11-08. 2014. URL: https://jdh.hamkins.org/transfinite-recursion-as-a-fundamental-principle-in-set-theory/.

[15] Karel Hrbáček and Tomáš Jech. *Introduction to set theory*. eng. Third edition, revised and expanded. Pure and applied mathematics. A series of monographs and textbooks ; 220. Boca Raton: Taylor & Francis, 1999. ISBN: 0-8247-7915-0.

[16] Akihiro Kanamori. *The higher infinite: large cardinals in set theory from their beginnings*. Springer, 2003.

[17] Asaf Karagila. *Lecture Notes: Large Cardinals*. 2025. URL: https://karagila.org/files/LC-2025.pdf (visited on 02/04/2026).

[18] Laurie Kirby and Jeff Paris. "Accessible independence results for Peano arithmetic". In: *Bulletin of the London Mathematical Society* 14.4 (1982), pp. 285–293. DOI: 10.1112/blms/14.4.285.

[19] Alberto Marcone and Antonio Montalbán. "The Veblen functions for computability theorists". In: *The Journal of symbolic logic* 76.2 (2011), pp. 575–602. URL: https://arxiv.org/pdf/0910.5442.

[20] Jeff Paris and Leo Harrington. "A mathematical incompleteness in Peano arithmetic". In: *Handbook of Mathematical Logic, edited by J. Barwise*. North-Holland, 1977, pp. 1133–1142.

[21] Michael Rathjen. "The art of ordinal analysis". In: *Proceedings of the International Congress of Mathematicians*. Vol. 2. European Mathematical Society. 2006, pp. 45–69.

[22] Stephen George Simpson. *Subsystems of second order arithmetic*. Vol. 1. Cambridge University Press, 2009.

[23] Will Sladek. *The termite and the tower: Goodstein sequences and provability in PA*. 2007. URL: https://andrescaicedo.wordpress.com/wp-content/uploads/2017/09/sladekgoodstein.pdf.

[24] Alan Mathison Turing et al. "On computable numbers, with an application to the Entscheidungsproblem". In: *J. of Math* 58.345-363 (1936), p. 5.

[25] Veritasium. *Math's Fundamental Flaw*. 2021. URL: https://www.youtube.com/watch?v=HeQX2HjkcNo (visited on 11/06/2025).

[26] Veritasium. *The Man Who Almost Broke Math (And Himself...) - Axiom of Choice*. 2025. URL: https://www.youtube.com/watch?v=_cr46G2K5Fo (visited on 11/08/2025).

[27] Vsauce. *How To Count Past Infinity*. 2016. URL: https://www.youtube.com/watch?v=SrU9YDoXE88 (visited on 11/04/2025).

[28] Googology Wiki. *Veblen function*. 2025. URL: https://googology.fandom.com/wiki/Veblen_function (visited on 11/12/2025).

[29] Wikipedia. *Hausdorff maximal principle — Wikipedia, The Free Encyclopedia*. 2025. URL: https://en.wikipedia.org/w/index.php?title=Hausdorff%5C%20maximal%5C%20principle&oldid=1300391387 (visited on 10/07/2025).